

More fuss about genetics and embryos

General alarm about genetics continues, fanned by the expectation that what may one day be possible will be a reality tomorrow. Moderation is the most urgent need, but understanding would also help.

THE danger that general excitement about the application of human genetics and embryology will get out of hand increases daily. The recent holiday season has accentuated the danger by emptying the newspapers of more substantial news and, perhaps, by giving people an opportunity to reflect on such developments as China's eugenic ambitions (see *Nature* 367, 1 & 3; 1994). The result is that people generally are being invited (by tabloid newspapers, for example) to contemplate that potential parents will soon be plotting the birth of 'designer' offspring, freed from genetic handicap by genetic screening as embryos and then nurtured in an appropriate uterus by *in vitro* fertilization (IVF). Sales of Aldous Huxley's *Brave New World* have no doubt picked up. In Britain, public excitement has been fanned by spurious publicity given to studies by a research group in Edinburgh of the maturation of ovarian follicles, and by the suggestion that human ova for IVF may soon be won from aborted human fetuses. (The publicity is spurious because the news is six months old, see *Nature* 354, 372; 1993).

There are several reasons why the excitement is misplaced, not least that the more outlandish possibilities are now regulated by the law. In Britain, for example, proposals to use fetal ova for IVF would have to be approved by the Human Fertilization and Embryology Authority, which embarked last week on a public consultation on the ethical issues involved. But that will remain an academic issue while so little is known of the function of the 100,000-fold excess of ova with which the female fetus is endowed. At least until it has been established that the ovarian follicles that ripen naturally are selected at random, and that the ova they contain are not in some way (metabolically or genetically) superior to those discarded, only a rash gynaecologist would wish fetal ova on infertile women patients.

Screening

Much the same is true of the other element in the popular recipe for the birth of designer offspring — the genetic screening of the embryo. The first demonstration of the use of genetic probes to determine single aspects of an embryonic genome are only now being published (see, for example, *Nature Genetics* 6, 19; January 1994). To suppose that these demonstrations will be quickly turned into a routine for screening human embryos for a whole battery of genetic attributes is to suppose a great deal too much, even if, in the long run, physicians come to recognize that it will be

anomalous that the time-consuming and expensive (in human resources) procedures of IVF should yield offspring with avoidable genetic diseases.

Difficulty

The underlying difficulty is that the spate of information now flowing from human genetics and embryology has engendered the illusion that everything is possible. (Triumphalist talk by some practitioners has not helped.) But everything is not yet possible, nor will it be. And, as with other developments in science that lead to new technology, the time-lag between the recognition of a technological possibility and its reality will always be substantial. Given the proper delicacy with which physicians working in these fields approach these issues (and the supervision of ethical committees), fears that whole societies will suddenly be overwhelmed by novel methods of reproduction are a needless self-induced nightmare.

A more substantial moderating influence on public excitement must be a more rounded and accurate view of what constitutes a gene. Despite the hectares of newsprint now devoted to explanation of the concept, semantic confusion persists. The foundation of molecular genetics is the notion that a gene is an identifiable stretch of genomic DNA, recognizable even from the nucleotide sequence from which RNA can be meaningfully transcribed, which may be complicated by irrelevant sections (introns) and which will have one or more end-points. But that, of course, is not the gene referred to in generally current phrases such as the 'gene for haemophilia'. The stretch of DNA eventually yields a protein called Factor VIII, which has a crucial role in normal blood-clotting. The 'gene for haemophilia' is a reference to one of several heritable variants of the Factor VIII sequence that yield a protein which is an inefficient component of the blood-clotting process.

Confusion

The semantic confusion may not matter much in the context of haemophilia, but is important when it afflicts general discussion of genes whose products may play several roles in human physiology. Even the notion that a 'gene for disease X' is a 'bad' gene in some absolute sense should long since have been exorcised from public discussion by the classic discovery (before molecular genetics came of age) that the version of haemoglobin that gives red blood cells the appearance of a crescent or a sickle causes anaemia in



homozygotes but protects against malaria those who inherit the sickle-cell variant of haemoglobin from only one parent. How common this phenomenon of heterozygous advantage may be remains anybody's guess.

Illustrations

Vivid illustrations of how the terms 'bad gene' and its opposite must be understood in only a relative sense are to be found in the current issue of *Nature Genetics* (6, 29; 1994), where a group from the Centre d'Etudes du Polymorphisme Humain (CEPH) in Paris describes the occurrence of known versions of the gene whose product is apolipoprotein E (called apoE), which has an important role in the regulation of blood cholesterol. There are three heritable versions of the protein, one of which is associated with ischaemic heart disease, and so is found to be less well represented among centenarians than among the general population. (The people with the 'bad' gene are supposed to have died from heart attacks, or alternatively from Alzheimer's disease, to which the same version of the apoE gene predisposes.) But another 'bad', or at least suspect, version of the same gene is found to be better represented among older than younger people, suggesting that it has a beneficial influence on longevity. The situation is further complicated by the known involvement of apoE in the maintenance of neurons. In other words, 'bad' and 'good' are strictly meaningless; their benefits may vary with the tissues in which they are active or with the stage of a person's development. What eugenicist would, in these circumstances, know which genes to get rid of?

It is a substantial slander of geneticists and their clinical associates that it should be so generally supposed that the subtleties of genetics and embryology are hidden from those parts of the research profession best placed to understand them. Indeed, the record of geneticists and embryologists in the past few years is thoroughly honourable. Who, after all, are those who have defined the questions on which morbid general interest centres if not geneticists and embryologists? That can be told from the recent report (see *Nature* 366, 498; 1993) on genetic screening by the Nuffield Council for Bioethics, which differs from most documents of this kind in providing a careful review of the genetic services now offered premaritally and during pregnancy in Britain. Diffidence about the social consequences of screening, and the effect on individuals of gloomy diagnoses, have so far limited screening to a mere handful of genetic conditions.

None of that implies that there are no difficulties ahead, as even China will discover. As knowledge of the often complex behaviour of normal gene products accumulates, premarital and prenatal screening will be more widely used, while greater numbers of people will learn something of their genetic constitution, and in some cases will have to live with uncomfortable knowledge of their likely fate. The immediate problems for society are not the questions of designer offspring now making the headlines, but those of the uses made of genetic information about individuals by outsiders. □

Charitable publication?

The principle that researchers shoulder responsibility for publication is under attack in Britain.

THE British institution called the Charity Commission is not particularly distinguished by assertiveness. Although the commission's responsibility is to determine whether the objectives of organizations qualify them for tax-exemption, and then to supervise their conduct of their affairs, battles about issues such as Scientology drag on for years, for example. But, out of character, the Charity Commissioners have now intervened in a matter concerning the responsibility of charitable grant-making organization for the publication of research they sponsor — and not on the side of the angels.

The origin of the commission's intervention is, in itself, a sad and tragic tale. In 1990, *The Lancet* published the result of a study of the survival of cancer patients at the Bristol Cancer Help Centre, which had previously claimed success in prolonging the lives of seriously ill patients by supplementing conventional therapy with 'holistic' alternatives, including diet and talk. The study concluded that, contrary to what had been claimed, the survival of the Bristol patients was shorter than that of controls. It soon emerged that the study was statistically flawed, whereupon one of the principal investigators, Professor Tim McElwain, committed suicide. The study had been funded by two British charities, the Cancer Research Campaign and the Imperial Cancer Research Fund (ICRF), most of whose work is carried out in its own laboratories.

The commission describes, in a statement last week, its investigation of the circumstances in which the research grant was made, evidently under pressure from supporters of holistic therapy persuaded that two orthodox medical charities had conspired to do them down. Concluding that there was nothing wrong with the manner in which the grant was made, but evidently embarrassed that the published product was mistaken, the commission concludes that "no one adequately supervised the study" and, worse, that the trustees of charities have a duty to arrange for the "evaluation of the products of research they have funded before the results are published". The commission says it plans to draw up guidelines on the subject.

This is a wrong-headed response to an acknowledged calamity. First, if there is a duty of supervision of principal investigators, it lies with their employers and not with those who support their research. Second, even employers are not well-placed to supervise the intellectual content of what researchers in their employ are about. Indeed, Professor Nick Wright, director of clinical research at ICRF, was correct last week to warn of the commercial or political bias that might thereby colour published research. The Charity Commission would be well advised to change its tack. Otherwise, it will quickly turn all medical research charities into in-house organizations, which is not what Britain needs. □

Academy under fire over plans for new study of DNA statistics...

Washington. Less than two years after publishing a highly influential report on the use of DNA technology in forensic science, the US National Academy of Sciences (NAS) is about to launch a new, fast-track study of the use of DNA evidence in court cases.

But proposals for the new study have infuriated scientists and lawyers who fear it could undermine the recommendations of the original report. The critics charge that the Federal Bureau of Investigations (FBI) used improper methods to press for the study, and to confine its scope to an examination of statistics.

The report has been requested by the FBI. The Academy will conduct the work through its research arm, the National Research Council (NRC), and says that the report is needed to take account of new data published since the original study.

William Sessions, the former FBI director, asked for the study last April, claiming in particular the need to incorporate new data. He requested that it focus in particular on the use of the so-called "ceiling principle" in the statistical evaluation of DNA evidence. This uses conservative 'guesstimates' to evaluate the probability of a DNA match in cases where the data does not exist to support a more accurate estimate.

The principle has been widely used in court cases since being proposed in the first academy study. Its intention was to reduce detailed discussion of uncertainties involved in very low probabilities. In practice, however, it has often been used by defence lawyers to convince courts that there is too great a chance of a false conviction in cases depending on DNA evidence, resulting in recurrent acquittals both in the United States and elsewhere.

Critics are concerned that the new study will issue revised rules on statistical evaluation that are more to the liking of prosecutors and the FBI, and that its exclusive focus on statistics will undermine other parts of the old report which the FBI wishes to ignore, such as the requirement for forensic laboratories to publish error rates after independent inspection.

If it proceeds, the study will be conducted by a panel chaired by population geneticist James Crow of the University of Wisconsin, and will take just over six months to complete. It will proceed only if the NRC can raise the \$300,000 needed to pay for it. So far, it has received more than half that amount from the National Institute of Justice (NIJ) — a sister agency of the FBI in the Department of Justice — and smaller contributions from the National Institutes of Health and the Department of Energy.

Critics argue that the personal request from Sessions led the NRC to overrule the advice of its own Commission on Life Sciences, and proceed with the study. They also cite a letter from John Hicks, assistant director of the FBI laboratories division, to the director of the National Institute of Justice, as evidence that NIJ funding for the study was conditional on its scope being restricted to statistics.

Richard Lewontin of Harvard University, a population geneticist whose work strongly influenced the first report, has written to Bruce Alberts, the president of the Academy, attacking the FBI's conduct and warning that "there is no way that the NAS/NRC can come out of this affair undamaged if it persists" with the report.

In his letter, Lewontin suggests that the new panel will either be dominated by scientists sympathetic to the FBI — in which case it will be seen as "rigged" — or it will be balanced "by others like me", ensuring "divisiveness, struggle and confusion for the courts and the scientific community."

Another population geneticist, Jerry Coyne of the University of Chicago, says he is "not convinced" that the new data justify a new study, and that he is "disturbed by the willingness of the National Academy to do the bidding of the FBI." Peter Neufeld, co-chair of the DNA Task Force at the National

Association of Criminal Defense Lawyers, says the Academy "has been compromised" by its decision to do the report.

But Bruce Alberts stands by the decision of the NRC under his predecessor, Frank Press, to go ahead, saying that new data have enabled people to dispute the original report's findings. "My feeling is that we have to go back and look at it again," he says. "I think we would have to do that independent of the FBI's view."

"There were plenty of checks and balances along the way here," says Eric Fischer, chair of the NRC's biology board and project director for the new report. The NRC has two reasons to proceed with the report, says Fischer: it was requested to do so by a federal agency — the FBI — and is required to respond; and there is "extra data and analysis" now published and available to do an update, denying that a focus on statistical evaluation will block out other issues.

But Eric Lander of the Whitehead Institute at Cambridge, Massachusetts, a member of the panel that produced the original report, says the update is only necessary because the administration failed to meet the panel's call for a "standing committee" to keep its recommendations up to date. "There's lots of new information and new evidence, but no mechanism for dealing with it," he says.

Colin Macilwain

... as confusion leads to retrial in UK

London. The Court of Appeal in London has ordered the retrial of a man convicted of rape after being identified through DNA evidence alone, because of confusion over the way in which the statistical interpretation of the DNA evidence was presented at his original trial.

The court ruled that evidence given by one of the forensic scientists, and the summing up of the judge, had fallen into the so-called 'prosecutor's fallacy' — the term used to describe confusion between two methods used to interpret the significance of an apparent match between two DNA samples.

The ruling may prompt a flood of similar appeals that DNA evidence had been wrongly interpreted to juries. Indeed, it is already

fuelling the heated debate in the United Kingdom and the United States (see above) about how to ensure the reliable use of DNA profiling techniques in the courtroom.

"This is the biggest straw that has been given to defence lawyers seeking to pick holes in the use of DNA evidence for a long time," says John Brookfield, a geneticist at the University of Nottingham. "I fear that this one will run and run."

The defendant, Andrew Deen, had been arrested during a random police check in Manchester, after his DNA profile was found to 'match' that of semen samples taken from a student who had been raped shortly beforehand. In February 1990, Deen was sentenced to 16 years in prison for this and two other rapes in the area.

Last summer, however, the appeals court agreed to hear evidence in Deen's defence, and in particular criticisms of the way in which DNA evidence was handled during his trial. The lawyer who prepared Deen's defence, Mike Mansfield, had previously obtained the release of the 'Birmingham ►

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

**Michael Mansfield:
showed up 'fallacy'.**

Six', partly on the grounds that the forensic evidence used to convict them for pub bombings in the 1970s was flawed.

Mansfield argued that although the ten bands used to compare Deen's DNA samples with those of the victim matched perfectly, the potential significance of discrepancies in two other bands had not been sufficiently explained to the jury.

He also argued that one of the prosecution's expert witnesses, as well as the judge, had confused two different sorts of probability. One is the probability that DNA from an individual selected at random from the population would match that of the semen taken from the rape victim, a calculation generally based solely on the frequency of different alleles in the population.

The other is the separate probability that a match between a suspect's DNA and that taken from the scene of a crime could have arisen simply by chance — in other words that the suspect is innocent despite the apparent match. This probability depends on the other factors that led to the suspect being identified as such in the first place.

During the trial, a forensic scientist gave the first probability in reply to a question about the second. Mansfield convinced the appeals court that the error was repeated by the judge in his summing up, and that this slip — widely recognized as a danger in any trial requiring the explanation of statistical arguments to a lay jury — justified a retrial.

In their judgement, the three appeal judges, headed by the Lord Chief Justice, Lord Farquharson, explicitly stated that their decision "should not be taken to indicate that DNA profiling is an unsafe source of evidence".

Nevertheless, with DNA techniques being increasingly used in court cases, some forensic scientists are worried that flaws in the presentation of their statistical significance could, as in the Deen case, undermine what might otherwise be a convincing demonstration of a suspect's guilt.

Some now argue, for example, that quantified statistical probabilities should be replaced, wherever possible, by a more descriptive presentation of the conclusions of their analysis. "The whole issue of statistics and DNA profiling has got rather out of hand," says one.

Others, however, say that the Deen case has been important in revealing the dangers inherent in the 'prosecutor's fallacy'. They argue that this suggests the need for more sophisticated calculation and careful presentation of statistical probabilities.

"The way that the prosecution's case has been presented in trials involving DNA-based identification has often been very unsatisfactory," says David Balding, lecturer in probability and statistics at Queen Mary and Westfield College in London. "Warnings about the prosecutor's fallacy should be made much more explicit. After this decision, people are going to have to be more careful."

David Dickson

France gives broader role to defence research panel

Paris. The French minister of defence, François Leotard, last week expanded the Scientific Council for Defence — the committee that advises the government on defence research — to include more representatives from universities, public research organizations and private industry.

Leotard described the main purpose of this move as being to make "tighter and more coordinated links" between civil and military research. But it is also being seen as a step towards a single integrated approach to technology policy, covering both the civilian and the military sector.

The head of the enlarged council will be André Giraud, a former head of the French Atomic Energy Commission (CEA). Giraud himself created the council in 1986 when he was minister of defence in the conservative government of Jacques Chirac.

Since then, its main activity has been to study specific areas of weapons technology at the request of ministers. In its expanded role, the council will monitor all advances in research for their potential impact not only on new weapons systems but also on other activities, including civilian industry.

This change highlights two trends. One is that the French military is turning to civilian science because, according to Leo-

weapons systems to defend its territory, to protecting its "image and international influence, including its economic strength and science base".

In the short term, it is also clear that France hopes to save some of the 104,000 jobs expected to disappear in the arms industry before 1997 by converting parts of the military-industrial complex to civilian goals.

Behind both trends are the implications of the end of the Cold War. One concern in France is that Europe will end up perched on a US-Russian axis. To avoid this, France seems likely to abandon its autonomy in weapons systems and support a European effort; a recent report to the government recommended that it share the cost of developing weapons (except nuclear) and other military technology with its European partners.

Observers in Paris feel that both initiatives are likely to propose increased funding for 'dual' high-technology research of both military and commercial interest. They may also lead to a reappraisal of the *grands programmes technologiques* in nuclear science, aeronautics and space, which still absorb much of France's research spending.

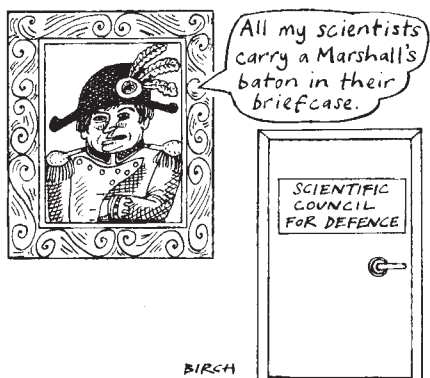
Quenzer says, for example, that although the *grands programmes* have been successful, the emphasis laid on them has led to neglect of other industrially important areas such as lasers and robotics. He says that a reorganization of the existing programmes is being discussed in order to produce greater industrial benefits, given that their political and strategic importance has been reduced by recent geopolitical changes.

The new council's first task will be to evaluate the importance of biology to defence. Quenzer says that France is concerned that developing countries could develop biological weapons, using advances in genetic engineering, for example, to produce cheaper and simpler weapons than nuclear bombs.

But the government's moves are not without their critics. François Clavier, a researcher at the CNRS' Institut de Physique Nucléaire at Orsay, and secretary of the Group of Scientists for Nuclear Disarmament, says he is disappointed that France has not grasped the opportunities offered by the end of the Cold War to boost civilian research.

Clavier says France should be reducing military spending, rather than increasing its weapons research effort and extending it into new areas such as biology. He is also worried that "dual" research may turn out to be a euphemism for "handing control of the civil science budget to the military".

Declan Butler



tard, it has belatedly recognized that, despite having given priority to its own research efforts in the past, France (like the rest of Europe) still lags behind the United States in fields such as space and computer technology and advanced avionics.

The French military can no longer afford to carry out such research on its own, particularly now that defence budgets are falling, and is therefore turning both to the civilian sector and to its European colleagues.

The other factor behind the decision to expand the scope of the council's work, says Alain Quenzer, its permanent secretary, is that France is broadening its concept of national defence beyond that of providing

German court blocks ban on animal tests

Munich. An appeals court in the state of Hessen has upheld a ruling by a local court that the state government infringed Germany's constitution last October when it tried to ban all experiments using animals from undergraduate courses at the University of Marburg (see *Nature* **365**, 778; 1993).

The country's animal protection law says that animals may be used in teaching only if no adequate alternative exists. But the local court had decided earlier in the month that this law could not be used to restrict the freedom of teachers, as guaranteed by the constitution, to choose how to teach.

In an apparent protest at this decision and support of a long campaign by students against the use of animals in undergraduate teaching courses, animal rights activists stole 86 rabbits and 78 dwarf hamsters being used in experiments at the university in a raid on Christmas night.

The students' campaign has become focused on a single experiment on anaesthetized rats which still remains in the zoology curriculum, and which they say could be replaced by a film. It was the state government's attempt to ban this experiment that has now been rejected by the courts.

The decision has increased the polarization between opposing sides in the dispute. The Health and Research Society (Gesellschaft Gesundheit und Forschung), which protects the interests of research workers, has described it as a "victory for science" that gives biology teachers more freedom to decide how to conduct classes.

But Brigid Völlm of the Bundesverband SATIS, a national student pressure group concerned with the misuse of animals in teaching, describes the decision, which was made unusually rapidly, as a "catastrophe"

for animal rights activists.

"It means that we will no longer be able to fight against animal experiments using the animal protection law," says Völlm. But she adds that her organization will continue to defend the rights of students not to work with animals, on the grounds that requiring students to do so infringes their own individual freedom.

The stolen animals have not yet been found. Gerald Heldmaier, a professor of

zoology at Marburg who claims that the number of animals used in his course has been reduced to a bare minimum in response to public concern, describes the theft as "criminal".

Heldmaier says that the lives of the animals have been put in danger. But Völlm denies this, saying that a veterinary surgeon was present during the raid, and that the animals are being well cared for in private homes.

Alison Abbott

Attacks prompt demand for new law

London. Britain's Home Secretary, Michael Howard, has been asked to introduce tougher legislation to curb attacks by animal rights activists on individuals and institutions involved in medical research.

The request has come from Sir Colin Berry, professor of morbid anatomy and dean elect of the London Hospital Medical College, and president of the Research Defence Society (RDS), a body set up to defend the use of animals in research.

It follows a recent upsurge in attacks by animal rights groups. These include letter bombs sent to the home of a prominent physiology researcher in Oxford, to employees of Glaxo's laboratories in Stevenage and to a biological organization in London.

RDS believes that, with 34 incidents in the past three months, the attacks herald a new protest campaign. "In terms of number of violent attacks, this is the most worrying escalation that we have since the animal rights extremist started operating in the mid-1970s," says Mark Matfield, the executive director of the society.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Rev. Features

Rough justice: can tighter legislation stop animal rights attacks such as this?

No group has claimed responsibility for the latest round of attacks. Some newspaper reports, however, have suggested that they may have been carried out by a previously unknown group calling itself the Justice Department.

RDS members point out that the government has already agreed to clamp down on the activities of hunt saboteurs by including a new crime of "aggravated trespass" in its criminal justice bill. This bill was presented to parliament shortly before Christmas and will soon be debated in the House of Commons.

The RDS proposal has been backed by various moderate groups concerned with animal welfare. "People are breaking the law, and that is not the way to reduce the use of animals in research" says Julia Fentem, scientific liaison officer of the Fund for the Replacement of Animals in Medical Experimentation.

Some medical scientists, however, fear that tough new legislation might further polarize the debate. They point out that existing legislation covers violent crimes committed by protesters, and are concerned that an over-reaction to recent events might provoke a backlash by hardening the attitude of animal rights groups.

"Long prison sentences may well give researchers some protection against the more fanatical protesters, but I do not think that tougher sentences will in themselves be any sort of deterrent," says Sir Dai Rees, the secretary of the Medical Research Council.

David Dickson

Anti-HIV claims were 'grossly distorted'

London. Claims by Britain's *Sunday Times* that the human immunodeficiency virus (HIV) does not lead to AIDS, and that the AIDS epidemic in African countries has been exaggerated, have been strongly denied by a doctor whose work has been used as a central component in the newspaper's justification of its case.

Angelo d'Agostino, chairman of a Nairobi hospice, has said that he wants to "categorically distance" himself from the "gross distortions" in a *Sunday Times* article about his work, published last August.

In that article, Neville Hodgkinson, the newspaper's science correspondent, reported that HIV-infected orphan children in d'Agostino's hospice were not in the process of developing AIDS, and that this experience "flies in the face of the conventional theories about the history of Aids in Africa".

D'Agostino, in a statement on 6 January to *The Independent on Sunday* — which has itself recently published features highlighting the AIDS epidemic in African countries — says that Hodgkinson did not indicate his own views during their three-hour meeting.

He said that he "cannot help but wonder about the motivation for such terrible journalism and, more so, decry the terrible effects on the unsuspecting public who are given a false sense of security and run the very real risk of contracting an incurable disease as a consequence".

D'Agostino asked the *Sunday Times* to correct Hodgkinson's article soon after it appeared, but received no acknowledgement. He issued a press release to draw attention to the erroneous report, and wrote again to Hodgkinson. No correction has been published.

Maxine Clarke

Mars probe lost through 'management faults'

Washington. The project management and systems engineering procedures used by the US National Aeronautics and Space Administration (NASA) on its Mars Observer mission have been scathingly attacked by two panels responsible for investigating the failure of the mission.

A last-minute change to the flight plan of the spacecraft probably led to the mission's demise last September, when contact was lost as it entered its orbit around the planet and prepared to release equipment on to its

surface (see *Nature* 365, 3; 1993).

But the *post mortem* on the mission's failure, made public in Washington last week, has pointed to more fundamental causes, cataloguing a pattern of loose management and non-existent systems engineering on the \$980-million project.

A report to NASA from an independent panel chaired by Timothy Coffey, research director at the Naval Research Laboratory in Washington, found the space agency's management of the programme failed to adapt as

the concept of the Mars Observer changed radically, from the original idea of several craft based on existing technologies to the complex spacecraft that was finally launched by the agency in 1992.

In effect, NASA developed a production-line system of standard missions based on cheap, readily available commercial satellite technology, and then used it to build a one-off interplanetary spacecraft, whose design was often heavily improvised. "The fundamental problem was they thought they had a routine mission, and they didn't," says John Pike, a space specialist with the Federation of American Scientists.

Coffey says that systematic weaknesses evolved in the programme as it developed over the years, and "the discipline and documentation culture was inappropriate for a one-of-a-kind mission". He adds that the fixed-price contract under which contractor Martin Marietta built the spacecraft was "appropriate for building things that you already know how to build".

The Coffey report finds that the most likely reason for the failure of the mission was the accidental mixing of a small quantity of fuel in the fuel-pressurization system, rupturing tubing and sending the craft into a slow spin.

If this did indeed happen, the spacecraft

Will NASA make up for lost data?

Washington. Scientists who lost instruments on the Mars Observer would like to see a re-flight of an identical spacecraft in 1996 to achieve the mission's scientific objectives. But according to officials at the US National Aeronautics and Space Administration (NASA), the agency is more likely to split the experiment payload between two or more smaller spacecraft — that is, if it can afford any new missions at all.

Ironically, the failure of the Mars Observer makes a re-flight the safest, and probably the cheapest, option for recovering global Mars data, according to Arden Albee, a project scientist at the California Institute of Technology.

Albee says that, because investigating committees from inside and outside NASA have gone over the spacecraft design in meticulous detail, the Mars Observer is much better understood than any new spacecraft would be. In addition, he says, spares for most of the main instrument components have already been procured.

Launching a single 'Mars Observer 2' is unlikely, however, given the embarrassing and conspicuous failure of the first spacecraft. Dan Goldin, NASA's administrator, is said to be against the idea, partly because of the way it would be perceived outside the agency.

One principal investigator suggests that re-flying the same spacecraft, even if it makes technical sense, would be "politically incorrect". A Mars Observer 2 mission would also require either a Shuttle flight (which are all booked) or a Titan rocket (which is prohibitively expensive).

The preferred option at NASA headquarters is therefore to fly most or all of the seven original Mars Observer science instruments on smaller spacecraft that could be launched by cheaper Delta rockets. These would probably be sent to Mars during successive launch 'windows', which fall every two years.

The first possible opportunity would be in 1996. Goldin has already ruled out a make-up mission this year, as the only

spacecraft that could have been ready in time would have been the Pentagon-developed Clementine (see *Nature* 363, 201; 1993), which would have yielded a minimal scientific return at Mars. "You would have been doing something just to do it," says Albee.

A team led by Charles Elachi, assistant director of NASA's Jet Propulsion Laboratory in Pasadena, California, spent last autumn looking at options to recover the data that were to have been collected by Mars Observer. This was designed to conduct a year-long global study of the planet's geology, surface chemistry and weather.

The team considered between 20 and 25 possibilities that could be ready for a 1996 launch. Most of them, however, involve adapting an Earth-orbiting spacecraft to fly to Mars, which presents another political problem, as that strategy was partly responsible for the Mars Observer failure (see story above).

Meanwhile, the scientists who lost experiments on Mars Observer are worried that if their instruments do not get selected for the first re-flight, the space agency will not have the funds for a second, and that even if it does, the original idea for the Mars Observer mission — a comprehensive, synoptic view of the planet — would be gone.

Indeed, many of the Mars Observer principal investigators, who lost a decade's work when the spacecraft disappeared in August, wonder if there will even be a 1996 launch. NASA already has a Mars mission on the books for that year — the MESUR (Mars Environmental Survey) Pathfinder, a small lander that will test technology for a future Mars surface network.

Advocates of the proposed Pluto Fast Flyby had also hoped for a new funding start in 1996 (see *Nature* 358, 701; 1992). With Congress and the White House warning that NASA can expect less money for space science next year than this year, the planetary community now has to contend with an unplanned mission elbowing its way into the funding queue.

Tony Reichhardt

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Great expectations: the Mars Observer prior to its launch.

would have been unable to charge its batteries, and will now be "dead" and drifting through space, having failed to enter its planned orbit around Mars.

The panel found that the mixing of fuel is likely to have taken place because some fuel evaporated and leaked through check valves that were supposed to contain it on the 11-month mission to Mars. Tests conducted for the panel showed that two table-spoonfuls of fuel would have leaked in this

way, and that this was easily enough to rupture the tubing.

But the check valves were never designed to prevent such vapour leakage. They were required to do so only because of a last-minute change in the flight plan which delayed pressurization of the fuel tanks until the probe reached Mars.

Peter Wilhelm, director of the Naval Center for Space Technology and a panel member, says the late change to the flight plan was intended to avoid another possible leakage problem with a regulator valve. "If you were going to do it again, you'd do it in a way that you did not have to contend with either," he says. "It is an overall system design problem. We found a whole bunch of problems as a result of a lack of systems engineering."

The panel also criticizes the mission planners for switching off communications with Mars Observer during the fateful manoeuvre, pointing out that this decision was related to budget constraints. The telemetry would have remained active only if NASA had been able to afford to test the ability of a key transmitter component to withstand the shock of the manoeuvre. But the complete loss of communications with the spacecraft means that the agency will probably never know what actually happened to it and caused the mission's failure.

A second report, prepared by a team at the Jet Propulsion Laboratory (JPL) in Pasadena, California — which was responsible for managing the contractor on NASA's behalf — is even more scathing about the systematic deficiencies of the mission.

The JPL report notes "a lack of top-down, system-engineering design approach to fault protection" and adds that the investigation team's own brief survey uncovered numerous design problems which had not been noted during mission development. It also says that proper reliability assurance was not done, and that documentation was often wrong.

One former senior NASA official says privately, however, that JPL gave only low priority to the Mars Observer mission, which he describes as "an experiment by NASA to get JPL to produce a cheaper mission". The mission, the former official says, "skidded over the launch line with desperate budget difficulties".

The Mars Observer experience is certain to have an impact on NASA's programme management strategy in future. "We will be examining these findings and developing a corrective action plan," says Wesley Huntress, the agency's assistant administrator.

In particular, questions will undoubtedly be raised about the administration's strategy of increasing the amount of freedom given to government contractors, with more fixed-price contracts and less close monitoring on a day-to-day basis. With Mars Observer, NASA appears to have taken this "hands-off" strategy a step too far.

Colin Macilwain

Polish science gets priority — but little extra funding

Munich. Poland's new government has proposed a small increase in this year's budget for science, which has suffered severe cuts since the collapse of communism. The increase falls well short of a promise made by in November by Waldemar Pawlak, the prime minister, to double the country's research budget in 1994; but it does reflect the government's decision to make science one of its priorities.

The Polish Scientific Research Committee (KBN), which functions as a research ministry, is hoping to persuade the government to give it more money later in the year. It is also encouraging research institutes to raise the notoriously low salaries of young researchers, many of whom have been leaving basic research for better paid careers.

In his inaugural address to the national Parliament, the Sejm, last November, Pawlak promised to increase support for science, and in particular to follow guidelines adopted by the previous government in July, proposing to increase the research budget from 9,000 billion zloty (US\$450 million) to around 20,000 billion zloty in 1994, subject to general budgetary constraints.

Such an increase would have raised the proportion of Poland's spending on research from 0.6 per cent to 1 per cent of its gross domestic product (GDP). But many doubted that such an increase was likely to materialize, given the depth of the country's economic problems.

In fact, next year's science budget will be virtually constant in real terms. The final budget proposals, which will be debated in Parliament within the next few weeks, offer science 11,460 billion zloty, only slightly higher than the predicted rate of inflation.



Waldemar Pawlak:
optimistic promises.

Nevertheless science was one of five sectors given a rise in real terms.

Witold Karczewski, the president of KBN and a neurophysiologist who has already experienced four different governments since the committee was set up in 1991, had been hoping for 16 billion zloty, raising the level of funding to 0.8 per cent of GDP.

But he still hopes that science will benefit from a budget adjustment promised for summer, and is asking for a 'top-up' of 4,000 billion zloty. The extra money will be used to support priority research areas identified by the committee, including new technologies such as biotechnology and materials science, heart disease and cancer.

Meanwhile Karczewski is planning to distribute much more of the science budget directly to research institutes, and to reduce the amount given to other ministries to provide additional support for research in their own institutes. He hopes this will put pressure on these ministries to support his request for more funds.

This redistribution will have a further benefit. Poland's research institutes now enjoy considerable autonomy, and are able to set their own salaries for researchers. Although salaries differ between institutes, they are invariably low. Young scientists are paid particularly badly, receiving between a third and a half of what they might expect in industry.

Karczewski, worried by an internal brain-drain of young scientists who are now staying in Poland but moving away from basic research, introduced a grant system three months ago under which young scientists can apply for an additional grant of up to 70 per cent of their institutional salary while working on their PhDs.

But he would prefer to see a general rise in salaries rather than relying on such *ad hoc* arrangements. His decision to give research institutes more control of their budgets is intended to encourage them to respond to pressure from employees to raise salaries.

Alison Abbott

Jury still out in case of Mongolian 'fossils'

Beijing. The case of the "false" Chinese fossils reported recently by the Chinese Academy of Sciences may not be closed after all. Last year, a committee of palaeontologists concluded that shell-like fragments discovered in Mongolia by scientists from the Shenyang Institute of Geology and Mineral Resources, and reported as coming from the pre-Cambrian era, were not fossils after all, but organic fragments formed during laboratory analysis (see *Nature*, 366, 499; 1993).

However, Su Yang-Zheng, head of the stratigraphy and palaeontology research section at the Shenyang institute, has now written to the academy saying that it is too early to decide whether the fossils in question are false. According to Yang-Zheng, whether the fossils are in fact merely chemical deposits "still remains a question that requires a great deal of scientific experiment". Since experts currently have differing opinions, he says, "no definitive conclusion can be made at present."

You Qin Li

Organic farms face dilemma over new gene technology

San Francisco. Organic farmers in the United States, who provide a small but important market for agricultural biotechnology companies, are divided over whether to accept the use of recombinant-DNA products.

The debate has intensified as the National Organic Standards Board (NOSB), the advisory panel that helps to set standards for US organic farming, prepares to discuss recommendations on how to treat recombinant plants and pesticides.

The NOSB is expected to recommend either a moratorium on the use of recombinant DNA, or a general ban that would still allow companies to apply to use individual products. But it could also prohibit recombinant DNA products entirely.

Urgency has been given to the board's deliberations by the fact that three pesticides incorporating genetically engineered bacteria have recently arrived on the market, and several new vegetables, including Calgene's 'vine-ripe' tomato and a virus-resistant squash, are awaiting regulatory approval.

At least one biotechnology company has been vigorously lobbying the board to allow use of its genetically engineered products. Jerry Caulder, chairman of Mycogen Corporation in San Diego, says that his company feels NOSB should give its approval both because of the importance of the market and because rejection could harm the image of biotechnology.

In the United States, organic farmers are generally suspicious of recombinant-DNA technology and its impact on the web of ecological relationships on which their farms rely. The Organic Food Production Association of North America, a trade group, has rejected the use of recombinant-DNA technology, saying that it is incompatible with

organic philosophy and that there are in any case natural substitutes available.

Others, however, believe that some recombinant products could provide useful tools consistent with their philosophy. These accept that recombinant-DNA technology should be seen as a synthetic process, some of whose products can be used in organic farming provided they pass various tests.

Some of these groups would like to accept Mycogen's products, which are based on genetically engineered killed bacteria toxic to crop-devouring insects. The unaltered bacterium, *Bacillus thuringiensis*, has been a staple of organic methods, and some growers believe a more powerful version could attract more farmers to organic practices.

But others are concerned that the widespread use of a genetically engineered bacterium could stimulate insect resistance and destroy a powerful tool for organic growers. And, they say, recombinant DNA techniques in general contradict the organic philosophy of seeking balance in the agricultural system, rather than relying on narrow technical solutions.

"We have made significant achievements in terms of how to produce food without chemical alternatives, and I don't think anyone wants to throw that out," says Margaret Clark, a member of the NOSB and organic produce manager for a supermarket near Seattle, Washington.

Some researchers believe that acceptance of some biotechnology products by the organic community would be an important breakthrough. Peggy Lemaux, a cooperative extension specialist for the University of California, Berkeley, says the organic industry could bring a valuable perspective to biotechnology research. **Sally Lehman**

Indian scientists get larger share of laboratory profits

New Delhi. Scientists working for India's Council of Scientific and Industrial Research (CSIR) are to receive a share of the profits made by their laboratories from sponsored research projects as an incentive to bring in more contract work from industry.

This follows the government's decision to reduce its support for CSIR. At present the council is required to earn one-third of its budgeted expenses from sources outside the government. Next year this figure may rise to 50 per cent.

The move by the CSIR governing body revives a practice that was abandoned in 1977 after scientists started fighting over their shares. This is not likely to happen again, according to S. K. Joshi, head of the CSIR, because of the guidelines the council has developed for profit-sharing.

Under the new scheme, 40 per cent of profits will be distributed among the principal contributors to the project and 20 per cent among the supporting staff. Five per cent will be put in a welfare fund and the remaining 35 per cent will go to the reserve fund of the laboratory concerned.

CSIR has 40 laboratories. Joshi says the new financial incentives should be strong enough for its scientists to go out and attract industry money for their laboratories. Before the government's introduction of economic liberalization, Indian industry had no alternative but to seek CSIR's help in developing new technologies. But since the removal of curbs on technology imports in 1991, the CSIR has had to face stiff foreign competition.

The profit-sharing scheme was proposed by a committee of senior CSIR scientists to win back Indian industry. In the 1970s, when the incentive scheme was in operation, CSIR laboratories sold 200 technologies each year to industry. This dropped to 25 in the 1980s. "If we must attract our industries we have to give some incentive to our scientists," says Joshi.

The scheme is an extension of a procedure whereby CSIR scientists are already permitted to work as consultants for private or public-sector agencies. They can keep two-thirds of their earnings, up to a maximum of £2,000 sterling. The rest goes to laboratory reserve funds.

CSIR earned £2 million through the consultancy activities of its scientists last year, but expects to get more by raising the upper limit of scientists' consultancy fees to £6,000.

Joshi expects to see some visible increase in funds from industries within a year of the revival of the profit-sharing scheme. Last year, CSIR laboratories spent £80 million, two-thirds of it on salaries.

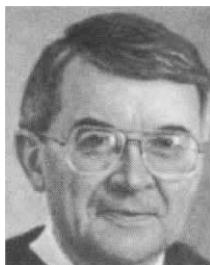
K. S. Jayaraman

New head promises shake-up at USGS

Washington. **Gordon Eaton (right), the director of Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York, has been nominated by President Bill Clinton as the twelfth director of the US Geological Survey (USGS).**

Providing he receives a quick confirmation from the Senate, Eaton should be in office by early spring. An Earth scientist specializing in tectonics, the 64-year-old Eaton is a former president of Iowa State University. From 1978 to 1981 he was associate chief geologist at the USGS.

Given recent tensions between the survey's three major programmes — in mapping, geology and water resources — Eaton says that one of his first priorities



will be to "knock the walls down, and to let people and funds flow across the boundaries between those divisions".

Eaton also stresses that "social relevance is going to

have to be a measure of a lot of the work that is done". For example, he says, the survey is likely to expand its research in areas of public interest such as water quality.

The result, he says, could be less emphasis on more traditional research topics such as hard-rock geology. **Tony Reichhardt**

Feuding hampers Internet's growth in Japan

Tokyo. The use of Internet, the global network of computer networks which links an estimated 20 million users worldwide, is about to take off in Japan. Two newly established companies are vying to provide commercial access to the network; and the government is about to upgrade the quality of the public sector portion of the Internet system, made up of computer networks linking universities and national institutes.

But rivalries between those responsible for competing networks in both the public and private sectors are still preventing the cooperation needed to ensure their maximum effectiveness. Such feuding over the control of Japan's computer networks is thus hindering their full development.

Japan has been slow to acknowledge the importance of computer networks, and the first high-capacity links to networks outside Japan were established through the initiatives of individual scientists rather than the government (see *Nature* 340, 670; 1989 and 356, 550; 1992).

The capacity of these links remains relatively low, and Internet traffic between Japan and the rest of the world is still limited in comparison with other advanced nations, being exceeded even by its small neighbour Taiwan (see caption above). But this situation is likely to change later in the year as the private sector and the government move to improve Japan's overseas computer links.

The first commercial provider of access to Internet has been InterCon International KK, a subsidiary of the US company InterCon Systems Corporation. This now offers full services with e-mail, file transfer and the ability to log into computers around the world. Another Japanese company, Internet Initiative Japan (IIJ), is expected to obtain a licence within the next month or so from the Ministry of Posts and Telecommunications to offer similar services.

But the start of commercial services has not been smooth. One of InterCon's first customers, a small Tokyo-based network

called TWICS backed by a non-profit educational foundation, was abruptly disconnected from its e-mail link to Tokyo University in November by the university's computer centre.

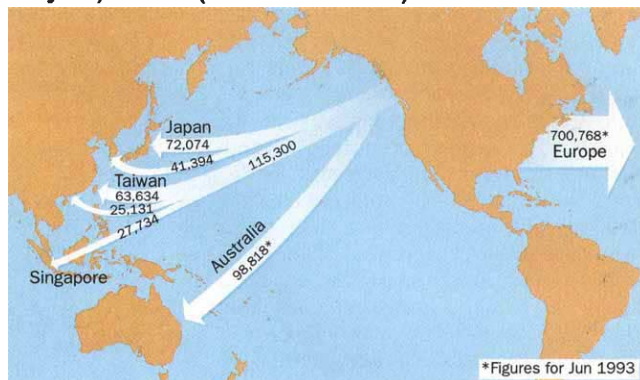
The disconnection took place shortly after TWICS signed up with InterCon, and inconvenienced several hundred TWICS users. TWICS was reconnected only after the intervention of Jun Murai of Keio University, the father of computer networks in Japan, and an adviser to IIJ.

Roger Boisvert, the president of InterCon, claimed that the "old boy network" of Tokyo University, IIJ and the managers of university networks in Japan were taking their revenge on TWICS for choosing InterCon rather than IIJ. The latter company has denied Boisvert's charges, and the university describes the incident as a "misunderstanding" over the timing of TWICS' move to InterCon.

But the affair has drawn attention to the competition that still exists between different networks. This year, the Science and Technology Agency (STA), with the support of various government ministries, will establish a new high-capacity interministerial network linking national institutes.

The network will have a high-capacity link to NSFNET in the United States, and a gateway to the Science Information Network (SINET) that links Japan's universities. It is run by the National Center for

NSFNET/INTERNET OUTBOUND TRAFFIC
Mbytes/month (November 1993)



Internet traffic between the United States and newly industrialized nations of the Western Pacific rim has surged in the past two years. The rise is particularly dramatic for Taiwan, with flow into and out of the National Science Foundation network NSFNET increasing from zero in September 1991 to 114,069 megabytes in November last year, exceeding the two-way traffic for its big neighbour Japan (102,566 megabytes). One explanation for the increase is the recent surge in the number of Chinese scientists returning to Taiwan from the United States. (Source: Internet Society)

Science Information Systems (NACSIS) of the Ministry of Education, Science and Culture.

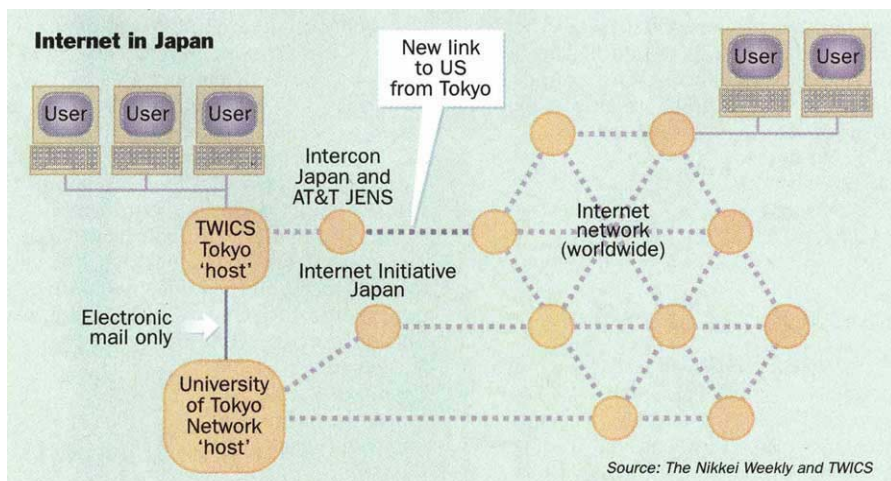
But as STA and NACSIS talk publicly of cooperation, each is manoeuvring to gain more control. STA plans to absorb into the new network the Todai (Tokyo University) International Science Network (TISN), established in 1989 on the initiative of individual Tokyo University researchers, which provides a link to the United States used by numerous universities and research institutes. According to one TISN user, however, NACSIS has been quietly lobbying TISN users to join its network.

Tateo Arimoto, director of STA's science and technology information division, hopes that eventually the various government-run networks will be merged into one "seamless" system. That dream still seems a long way off.

At present, at least six government-related networks have links to overseas: TISN, SINET, the Widely Integrated Distributed Environment (WIDE) established by Jun Murai, the Real World Computing project run by the Ministry of International Trade and Industry, STA's Japan Information Centre of Science and Technology, and the Japanese section of the BITNET network, run by Tokyo Science University.

Many observers feel that these organizations should pool their resources and buy one very-high-capacity international link. According to Arimoto, the US National Science Foundation (NSF) has been strongly urging Japan to do just that. But such cooperation still seems a long way off.

David Swinbanks



Validity of whaling data

SIR — For many years, scientists have been arguing about the validity of catch data from commercial whaling operations. Some parts of previously secret records from Soviet whaling in the Antarctic were recently made public during the plenary session of the Conservation Status of Marine Mammals at the Society of Marine Mammology's Tenth Biennial Conference in Galveston, Texas, on 12 November 1993. Actual Soviet catch data on right, humpback and blue whales from the 1960s were reported. These data were from one of the four factory ships that operated in the Southern Hemisphere after the Second World War. The catches for right, humpback and blue whales were reported as 717, 7,207 and 1,433 respectively. These numbers are much higher than were previously reported to the International Whaling Commission (IWC). The catches for humpback and blue whales were originally reported as 152 and 156 respectively.

Right whales have been protected under IWC regulations, and some earlier agreements, since the 1930s. However, Best¹ in 1988 reviewed three episodes of illegal exploitation by Soviet whaling fleets operating around Tristan da Cunha in the South Atlantic. These catches have been confirmed by the data presented at Galveston.

There have been constant rumours about illegal large-scale Soviet whaling operations not only in the Antarctic and South Atlantic, but also in the South and North Pacific. When I studied cetacean morphology at the land whaling station on Paramuschir Island (in the Northern Kurile Islands) at the end of the 1950s, I received anatomical materials not only from humpback but also from right whales.

It was also known that in the 1960s a Soviet factory ship illegally operated for a couple of weeks in the Okhotsk Sea and caught several hundred right whales. It was also well known in the Soviet Union that blue whales continued to be killed after they were protected by the IWC².

During the Galveston conference, I pointed out in my talk about the IWC that data problems exist with catch records from other countries as well. Kasuya³ and others have reported that catches of sperm whales were under-reported by both number and sex in Japanese land-based sperm whaling operations.

In these circumstances, it has been impossible to conduct a meaningful comprehensive review of the impact of all past commercial whaling operations. In order better to understand how various species of whales have been over-exploited, it would be highly desirable to investigate all whaling records now available. Such a

review would also help us to understand the potential for recovery of all depleted whale populations.

Alexey V. Yablokov

(Special Adviser to the President of Russia for Ecology and Health)
Kremlin, Moscow, Russia

1. Best, P. B. *Biol. Conserv.* **46**, 23–51 (1988).
2. Zemsky, V. A. & Sazhinov, E. G. in *Marine Mammals: Collected Papers* (ed Arsen'ev V. A.) 53–70. (All-Union Research Institute of Marine Fisheries and Oceanography, VNIRO, Moscow, 1982) (in Russian with English summary).
3. Kasuya, T. *Mar. Mammal Sci.* **7**, 230–257 (1991).

Patent medicines

SIR — Your article ("UK clinical geneticists ask for ban on the patenting of human genes", *Nature* **366**, 391; 1993) makes it clear that the mechanism of patent protection does not work very well when applied to therapeutic products that might arise from the knowledge of gene sequences. As the patent is a creation of human rather than natural law, the answer is surely to devise a new form of intellectual property that would serve the required purposes better.

The most important function of patenting pharmaceuticals is now to protect a company's investment in development costs. The trouble is that, unlike the situation in, for example, mechanical engineering, the protection arises (or fails to arise) before the bulk of these costs is incurred. This means that companies cannot afford to develop drugs whose patentability has been compromised by publication. It also means that holders of some fundamental patents reap rewards out of all proportion to the size of their own real investment.

However, and also unlike the products of mechanical engineering, the sale of pharmaceuticals is controlled by state licensing. Indeed, most development costs of drugs are incurred in order to comply with licensing requirements. I suggest that this investment would be fully protected if the right to quote the results of clinical trials and so on in support of licensing applications became a distinct form of intellectual property.

A. F. W. Coulson

*12 Eskside West,
Musselburgh,
Midlothian EH21 6PL, UK*

Referee bias

SIR — The controversial reviewing of a paper by expert referees may reflect a psychological built-in bias that characterizes any human judgement. With papers that, in referees' opinions, do not lack

adequate experimental and/or theoretical support, it appears that the driving force for their evaluation is their personal feeling.

Many researchers therefore feel aggrieved, because it is not clear what 'scientific' factors hamper publication. A possible answer lies in the fact that a 'controversial' paper raises a point that is source of a debate between opposite 'schools of thought'. Nevertheless, articles from different 'schools of thought' continue to be published, but the authors are seldom unknown. It may be that editors and/or referees are more favourably disposed to the publication of a controversial paper if its authors are authorities in the field. On this basis, authors lacking a 'pedigree' do not seem to be allowed to enter a scientific debate, given that fate has assigned the reviewing process to contrasting referees. We could call this behaviour 'Carneades' syndrome'. Carneades was a minor Greek philosopher, whose thought, even if of some importance, was never enough appreciated. This was underlined in the novel *The Betrothed* by Alessandro Manzoni, an outstanding Italian author from the Romantic Age, with the well-known question "Carneades, and who on Earth was he?"

Anonymity for both authors and reviewers could perhaps limit the risks of this syndrome, even if it is difficult to remove traces of an author's identity from a paper. We feel that new proposals about a reviewing process as devoid as possible of unconscious psychological influence could come from a public debate on this matter.

Giovanni Colonna

Raffaele Ragone

*Department of Biochemistry
and Biophysics,
2nd University of Naples,
via Constantinopoli 16,
80138 Naples, Italy*

Older than that

SIR — Punctuated equilibrium may have come of age (*Nature* **366**, 223–227; 1993) but its age is not 21 but at least 128. It was developed by P. Trémaux in 1865 in *Origine et Transformations de l'Homme et des Autres Êtres* (Hachette, Paris), as I indicated in 1989 in *Speciation and its Consequences* (eds D. Otte & J. A. Erdler, Sinauer, Sunderland, Massachusetts). The wonder is that this continental recipe of the mid-nineteenth century should today be relished as innovation of modern evolutionary theory.

G. Nelson

*School of Botany,
University of Melbourne,
Parkville, Victoria 3052,
Australia*

AIDS and *The Sunday Times*

SIR — I was astounded by your two-page leading article attacking *The Sunday Times* for having the temerity to publish a series of articles that run counter to the accepted theory of the causative agent of AIDS. I am a reader of both *The Sunday Times* and *Nature* (and a microbiologist), so feel qualified to give a dispassionate view.

In my opinion, it is quite probable that HIV infection is an important factor that leads to AIDS. However, this does not excuse the behaviour of a respected scientific journal in devoting precious space to attacking a leading exponent of a different view. To say that the "public interest requires that *The Sunday Times* should not follow its perverse line on the causation of AIDS" is outrageous. There are too many examples in the scientific literature of the then-accepted scientific dogma being overturned by a few individuals with the courage and intelligence to question it.

Surely the point of *The Sunday Times* articles is not to discourage 'safe sex' by teenagers, but to discourage a completely blinkered scientific approach that is 100 per cent certain that there is no other possible explanation for AIDS than HIV.

Richard James

University of East Anglia,
Norwich NR4 7TJ, UK

■ The following letter was submitted to *The Sunday Times* on 3 June 1993 but was not published.

SIR — The front-page piece about Wellcome ("Fears over the drug giant's funding of Aids research", 30 May 1993) muddles the very clear separation of the pharmaceutical company (Wellcome Foundation Ltd) and the charity (The Wellcome Trust). Even though your report says there is no suggestion that the trust has distorted its funding to benefit the company, it is full of implications that confuse the two, creating unjustified criticism of the trust.

In my research on components of the AIDS virus, I received useful assistance, in the form of experimental material, but no money or research resources from the pharmaceutical company — this research was funded by the Medical Research Council's AIDS-Directed Programme. I have also regularly acted as a referee on research proposals made to the Wellcome Trust, but I have never received any grant from it. Therefore I believe my comments are unbiased. I have personally observed the great emphasis in the trust on distancing itself from the pharmaceutical company.

When AIDS first became a disease of concern in the United States, the pharmaceutical company was in the enviable position of owning the only drug (AZT)

known to affect the progress of the disease and already approved for human use. Eight years later, it is now probable that although AZT changes the early development of symptoms, it has no favourable effect on the long-term prognosis. It was always known that it had harmful side-effects, but its long-term success in treating the disease could not have been assessed until recently.

It is wrong to criticize the trust for supporting research on AIDS. Other medical research agencies in the United States and the United Kingdom have rightly assigned large resources to the study and treatment of this important epidemic. As the largest source of medical research funding in the United Kingdom, the trust has a clear obligation to take part in this effort. Like other similar organizations (including the research councils which dispense government research funds), it invites scientists contemplating large research proposals to discuss them before making a formal application: but such applications are always evaluated by the traditional method of peer review and decided by an expert committee.

Research necessarily deals with uncertainty, and is undertaken because knowledge is lacking. To criticize trust support for Professor Anderson's epidemiological studies of the spread of the disease, or Professor Pinching's AZT trials, because they are alleged to have helped the sales of AZT, defies logic and betrays ignorance about the way medicine advances. If Mr Martin Walker believes the Wellcome Trust has used its power of patronage "to open doors for the company at the highest level" (in the words of his article), he will have to prove it in the face of strong evidence to the contrary.

David Blow

Imperial College of Science,
Technology and Medicine,
London SW7 2BZ, UK

■ The following letter was written to *The Sunday Times* in September 1993, but was not published.

SIR — I am writing to ask you to consider the manner in which your newspaper is treating the issue of HIV and AIDS.

In response to the 29 August article by Neville Hodgkinson, I wrote to your letters section. I received an acknowledgement and an apology that there was insufficient space to publish. Naturally, you cannot print all the letters you receive, but I was surprised that last Sunday's edition gave equal weight to correspondence supporting and deploring Hodgkinson's article.

I cannot believe that the balance of mail was reflected by the selection of letters you published. Is it your policy to back up

your own journalist, regardless of how misleading and irresponsible his article may have been? If so, I would ask you to give serious consideration to the consequences of this action.

All indicators suggest that the incidence of HIV infection in the United Kingdom is relatively low at present, but its high prevalence elsewhere in the world should guard us against complacency. You will be aware of the government money that has already been spent to inform the public on this issue, and that Hodgkinson is undermining this message. Can this really be justified as a public service, or even a legitimate use of press freedom? Scientists and doctors involved with AIDS know just how nonsensical Hodgkinson's articles are, but are powerless to dispel the confusion he has created among those less aware of the facts.

As a personal acquaintance of the original "HIV doesn't cause AIDS" guru (Dr Peter Duesberg), I know all the false premises and bogus arguments only too well. The arguments are a rerun of the tobacco companies' favourite old chestnut "smoking doesn't cause cancer", are equally futile and potentially just as damaging to public health.

Is it your intention to seek advice from more reputable sources and to try to set the record straight in future *Sunday Times* articles on this important issue?

James C. Neil

(Member of MRC AIDS Directed Programme Steering Committee),
Department of Veterinary Pathology,
University of Glasgow,
Glasgow G61 1QH, UK

■ The following letter was submitted to *The Sunday Times* on 14 December 1993, but was not published.

SIR — Following last Sunday's edition of *The Sunday Times* (12 December 1993), I feel compelled to write to you about the bizarre stance adopted by your Scientific Correspondent, Neville Hodgkinson, on the subject of HIV and AIDS.

Over the past 15 years I have been involved in several contentious issues which, on occasion, have brought me into conflict with the scientific establishment and members of my own profession. During this time I have always based my case on the scientific data available and, to this end, have organized conferences and edited books so that the data available can be properly evaluated. In addition, I have always invited co-editors who are regarded as pre-eminent in their field.

Many of the contentious issues in these different fields (biological effects of low-level lead, ionizing radiation, ozone depletion, global warming) have been resolved through communication, and common ground has been found between campaigners and the established experts.

I am sad to say that none of this applies

to the extraordinary campaign being conducted by your science correspondent. Not only are his arguments specious but he has singularly failed to address the large volume of evidence that undermines his point of view. By ignoring the scientific method — the scientific procedures laid down to distinguish fact from fantasy — he has not only discredited himself, he has seriously damaged the reputation of your newspaper.

Furthermore, your hostile attitude to John Maddox is both unfair and irresponsible. The reason, I suspect, that certain scientists dislike his approach is because he applies the scientific method with considerable rigour, as Jacques Benveniste found to his cost.

In my dealings with Maddox, I have never found him to be anything other than intellectually honest and courteous, and I think his response to your reporting of the AIDS issue is entirely appropriate. Far from being a triumph for *The Sunday Times* akin to your thalidomide campaign, I think your perception of the AIDS issue is akin to the Hitler diaries.

R. Russell Jones

Royal Postgraduate Medical School,
Hammersmith Hospital,
Du Cane Road, London W12 0NN, UK

Nuclear Ukraine

SIR — Your article (*Nature* 365, 599; 1993) on the US-Ukraine stalemate over nuclear weapons prompts the following remarks. The United States made a mistake in not recognizing Ukraine as a legitimate successor state to the Soviet nuclear arsenal and is still insisting that Ukraine give up its nuclear weapons to Russia. That policy goes back to President George Bush's ill-advised lecture to the Ukrainian parliament advising it not to leave the Soviet Union. John J. Mearsheimer (*Foreign Affairs*, Summer 1993, p. 50) has suggested that it is not too late to change course and to support a Ukrainian nuclear deterrent against Russia.

Ukraine, the Baltic republics and other former Soviet subject nations are currently targets of Russian neo-imperialist policies. Russia's immediate objectives are aimed at controlling territories necessary for the projection of its naval power. In the Far East, this has meant control of the Kurile Islands. Russia is willing to do without billions of dollars of aid from Japan just to prevent its nuclear submarines from being bottled up in the Sea of Okhotsk. The coastal real estate of the Baltic states is valuable to Russia for the same reason, and Russia has done all it can to put former communists in power there. They succeeded in Lithuania, but opponents of former communists won the elections in Estonia and Latvia. Russia then started a disinformation campaign

against Estonia and Latvia, with Boris Yeltsin himself complaining of "... a massive violation of human rights..." of their Russian minorities and halting the withdrawal of the Red Army to "protect" these colonists from the Soviet era. The *New York Times* (22 November 1992) subsequently reported that "a week-long investigation into Mr Yeltsin's allegations in Estonia and Latvia finds little evidence of human-rights abuses, at least as commonly understood".

The pressure continues, however, on the diplomatic front and has more recently taken the form of opposing the admission of the Baltic states to the European Union and of former Warsaw Pact members to NATO. The Cold War may be over, but Russian generals are all former Soviet generals, they still have a huge army, their military/industrial complex is still a drag on the economy and is yet to be dismantled, a sanitized KGB is still in business under a new name, and nobody has yet been punished. Can anybody be surprised?

Arno Arrak

5 Chatham Place,
Dix Hills, New York 11746, USA

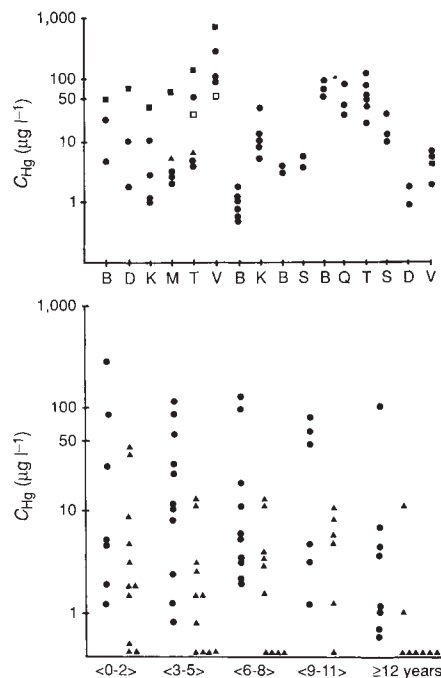
Mercury exposure

SIR — We would like to add to the accounts^{1,2} of mercury poisoning that have recently appeared. We discovered an elevated mercury concentration in the urine of a 2-year-old girl in a refugee camp in Germany. Subsequent investigations revealed that other family members and other inhabitants of the camp had high concentrations of mercury in the urine.

The camp is next to a coal-fired power plant and is inhabited mostly by refugees from former Yugoslavia and from Romania. Some families, including those with high mercury levels, have lived there for several years, and many of the younger children were born in Germany. Mercury concentrations were not always elevated. There were families with high levels and apparently unexposed families. Women tended to have a much higher concentrations (see figure).

After a long search for the source of this contamination, a physician from Kosovo helping us as an interpreter discovered that all the exposed families used a cosmetic bleaching ointment containing rose oil and a mercury salt; the ointment had to be buried for a year for proper seasoning. It was sold by a Turkish vendor. In two samples we found concentrations between 708 and 17,200 $\mu\text{g Hg per g ointment}$.

We went into two camps and investigated all children in whom we had detected elevated mercury values. Their age and urinary and blood mercury concentrations at the time of clinical investigation are shown in the figure. None had any



Top panel, urinary mercury excretion by individual members of 16 different families. (The letters represent the initial letter of the family name.) ■, Mothers; ▲, fathers; ●, children; □, female relatives. Lower panel, mercury concentration in urine (●) and blood (▲) in children at the time of clinical investigation. In some cases it was not possible to collect both samples.

dermatological, neurological or behavioural signs of acrodynia.

One woman with a Hg concentration of 150 $\mu\text{g l}^{-1}$ in urine and 29.9 $\mu\text{g l}^{-1}$ in blood had given birth to a normal baby (50 cm, 3,060 g, 34 cm head circumference). The placenta contained 120 $\mu\text{g Hg kg}^{-1}$. The mercury concentration in the umbilical blood was 8.8 $\mu\text{g l}^{-1}$. This boy was four weeks old when we saw him. He was clinically and neurodevelopmentally normal.

There are few data on the incidence of acrodynia at various mercury concentrations, some dating back 40 years³. In our patients, much higher mercury concentrations were measured and tolerated without the manifestation of clinical signs in several cases described in the literature. There may be a wide range of individual, possibly also ethnically conditioned, variation of susceptibility in children to low, chronic mercury exposure.

M. Otto

C. Ahlemeyer

Hedwig Tasche

K. E. von Mühlendahl

Akademie für Kinderheilkunde
und Jugendmedizin,
Ibarger Str. 200,
49082 Osnabrück,
Germany

1. Wendroff, A. P. *Nature* 347, 623 (1990).

2. Mills, A. *Nature* 346, 615 (1990).

3. Warkany, & J. Hubbard, D. M. *Lancet* i, 829-830 (1948).

The genetic code by numbers

The application of the theory of mathematical groups to the origin of the genetic code will startle molecular biologists, but is best regarded as a valuable exercise in classification.

THE problem of the genetic code has several facets, of which the most compelling is that of why it is just what it is. Once Crick and Brenner had established that the code is a triplet code, with each consecutive triplet in a molecule of either DNA or RNA corresponding to a single amino acid in the ultimate protein (*Nature* **192**, 1227; 1961), it was natural that people should look for an explanation, both for its own sake and because an understanding of how the code evolved must certainly be a pointer to the origin of life itself. It was already clear that the genetic code is not merely an abstraction, but also the embodiment of life's mechanisms; the consecutive triplets of nucleotides in DNA (called *codons*) are inherited but they also guide the construction of proteins.

So it is disappointing, but not surprising, that the origin of the genetic code is still as obscure as the origin of life itself. That is not to say that people have not been worrying about the problem, although most of the worrying seems to have been done at an early stage, in the 1960s and early 1970s, soon after the decryption had first been done. Is that a mark of people's disappointment, or merely a sign that they have had better things to do?

Historically, there have been two views of the origin of the code. The most obvious, that there must be some physical chemical interaction between a nucleotide triplet of the code in a nucleic acid polymer represented by a molecule of messenger-RNA (mRNA) and the amino acid it specifies in the eventual protein, was first advocated by Carl R. Woese and his associates. That was knocked down by the eventual demonstration that amino acid molecules are scavenged from the cytoplasm of a cell, and then carried to the ribosomes at which protein molecules are assembled from amino acids, by RNA molecules called transfer-RNA, tRNA for short. Although tRNA molecules are small as RNA molecules go, there is no obvious way in which the amino acid molecules they carry can interact with their signatures, the *anticodons*.

A little definition will help. The DNA code has four elements, called T, C, A and G (for thymine, cytosine, adenine and guanine). The RNA is similarly specified except that U (for uracil) replaces T. In double-stranded DNA, A and T are paired together in oppositely directed strands, as are G and C. In RNA molecules, tRNA molecules in particular, physical pairing of this kind is as often internal, helping to shape the secondary structure of the molecule. The character-

istic anticodon of a tRNA molecule is the ribonucleotide triplet that is complementary to the triplet of the mRNA specifying a particular amino acid. For example, if UGC is an RNA codon (which happens to specify the amino acid cystine), the anticodon will be ACG, invariably found unencumbered by internal pairing. The cystine molecule is then attached to the other end of the tRNA.

Long before that was known, some general features of the genetic code were plain. For one thing, there are $4 \times 4 \times 4 = 64$ possible codons, but only 20 amino acids are known naturally to occur. Even allowing for the need to specify the points at which the transcription of DNA into RNA normally stops, the genetic code must be redundant; some amino acids must be specified by many different codons. As early as 1961, Crick and Brenner put their money on degeneracy (rather than on the alternative that the superfluous codons meant nonsense). Among other things, they saw it as a way of explaining how microorganisms whose nucleotide composition is very different could have proteins with similar amino acid composition.

By 1968, Crick had carried the argument a step further (*J. molec. Biol.* **33**, 367; 1968) by demonstrating with cartesian logic that if the genetic code is now a triplet code, it must always have been a triplet code: a change in the base would always have meant the mistranslation of essential proteins and thus a loss, perhaps catastrophic, of darwinian fitness. Crick also put his weight behind the notion that the primitive genetic code would have been less specific than that which now holds sway, either specifying fewer amino acids or specifying classes of amino acids, perhaps defined by their acidity.

Much less has happened since 1968 than might reasonably have been expected. Perhaps the landmark is the paper by T. H. Jukes in 1983 (*J. molec. Evol.* **19**, 219; 1983) in which he used a comparison of the universal genetic code (applicable in all cells) and that in mitochondria (then just recognised to be different) to argue that the primaevial genetic code must have used sixteen anticodons to code for fifteen amino acids (with the extra anticodon used as a stop signal). That does at least point out that, when more is known of the dynamics of the genome, it should be possible to unravel the evolution of the code from the regulation and placing of tRNA genes and from the properties of defective pseudo-genes.

Meanwhile, quite a different way of regarding the evolution of the genetic code

has come to light, one likely to be more familiar to particle physicists than to molecular biologists. In the last issue of *Physical Review Letters* (**71**, 4401; 1983) for 1993, José Eduardo M. and Yvone M. M. Hornos, both from the University of São Paulo, argue that the problem of the genetic code is simply a problem in symmetry breaking similar to the reasons for the supposed existence of, say, the Higgs boson, and therefore best described by group theory.

The argument is not negligible. The starting point is the recognition that there are 64 different codons (or anticodons). If all the amino acids were once the same (or if the differences between them did not matter), that would be a highly symmetrical circumstance. But if differences matter, then the symmetry is destroyed or broken. And then it is permissible to look for chains of abstract mathematical groups which, if multiplied together so that the elements of the product group combine in ways determined by combinations of each component group, will reflect the classifications observed in the real genetic code.

The simplest analogy is with, say, the allowable electron states in the second principal quantum level of an atom — the line of the periodic table beginning with lithium. There are four of them, meaning that the electron shell concerned can accommodate eight electrons when allowance is made for the two-valued spin of the electron. But ordinarily, the existence of four states is not apparent. Only when there is a magnetic field does it become plain that one of these states has zero angular momentum, that the other three have unit angular momentum and that the latter three may be oriented in three different ways with respect to the external magnetic field. The textbooks are full of these, the nearest things to cladograms in atomic physics. They show how energy levels split into sub-levels under the influence of asymmetry.

So, it seems, it may be with the genetic code. The Hornoses have found a mathematical group with a 16-dimensional representation (the symmetric primaevial genetic code) which can be broken down into a product of simpler groups reflecting the pattern of redundancies observed. They make the system work, and even conclude that there may have been an intermediate code specifying only 15 amino acids, just as Jukes suggested. But sadly, even if the argument were much more compelling, only laboratory work can settle what really happened. That remains a challenge.

John Maddox

A question of making it work

Anders Björklund

SEVERE damage to the spinal cord has devastating consequences because the severed neural pathways do not regenerate. Two papers published elsewhere in this issue^{1,2} describe promising experimental results which demonstrate ways in which such damage might be at least partly repaired (although much of course remains to be done before therapies are likely to succeed in humans).

The execution of coordinated movements depends on control exerted by higher brain centres. After a transecting lesion of the spinal cord, the networks responsible for the generation of locomotion, which are located in the cervical and lumbar enlargements of the cord, are left intact but cannot be used because of the lack of connections with the supraspinal control pathways. The big challenge in spinal cord regeneration research is to find ways to stimulate axonal regrowth, and to modify the properties of the axonal growth trajectory in the damaged cord, in order to re-establish enough functional supraspinal connections with the isolated spinal cord segments. On the evidence of the papers by Schnell *et al.*¹ and Iwashita *et al.*² it appears that these strategies are starting to pay off.

In the mammalian central nervous system (CNS), severed axons exhibit a normal sprouting response but the sprouts fail to cross even narrow gaps and cannot grow back to their denervated targets. The failure of the axons to regenerate over long distances is clearly not due to an inability of the neurons to regenerate, but rather to the non-permissive nature of the CNS tissue environment: CNS neurons regenerate well, for example within a peripheral nerve, but the regrowing axons fail to penetrate mature brain or spinal cord tissue by more than a few millimetres (refs 3–7). Successful, directed axonal growth over long distances depends on interaction and balance between growth-promoting and growth-inhibiting factors, and it seems that regenerative failure is the result of a combination of active growth inhibition, exerted by the CNS tissue environment and the developing glial scar, and an inadequate supply of growth-stimulating factors necessary to sustain active regeneration.

An approach pursued by Schnell and Schwab¹ is to pave the way for regrowing axons by blocking a subset of inhibitory myelin-associated proteins with neutralizing antibodies. Axonal regeneration after partial cord transection (as observed in the cortico-spinal tract fibres, which express a particularly feeble sprouting response) is greatly increased after such antibody treatment. But even in the best

cases the number of successfully regenerating axons has been small.

Schnell *et al.*¹ searched for a neurotrophic factor which, together with blockade of growth inhibition, could increase the magnitude of cortico-spinal tract regeneration past a spinal hemisection lesion. The prototypic neurotrophin, nerve growth factor (NGF), stimulates regeneration after lesions in the NGF-sensitive peripheral sympathetic and sensory neurons^{8,9}, and in the brain a similar effect on axonal regeneration has been reported for the basal forebrain cholinergic neurons (the principal neuronal targets for NGF in the CNS)^{10,11}.

Schnell *et al.*¹ now report that another member of the neurotrophin family, neurotrophin-3 (NT-3), can enhance sprouting of the corticospinal axons, both during development and after axotomy in young adult rats. Although NT-3, given as a single injection at the site of injury, did not by itself promote axonal elongation, the authors observed some cases of regeneration over remarkably long distances down the cord when the NT-3 treatment was combined with inhibition-blocking antibodies. In the successful cases 5–10 per cent of the cortico-spinal fibres were estimated to have regenerated, and the maximum growth distances along the cord (15–20 mm) were of an order of magnitude that has previously been achieved only with grafted fetal neurons. So NT-3 may be a long sought-after candidate growth factor for one of the principal projection systems of spinal cord. NT-3 is indeed normally expressed in the spinal cord during fetal and early postnatal development, and *in situ* hybridization data indicate that its preferred high-affinity receptor, trkC, may be located on the cortico-spinal neurons¹².

A shortcoming of the Schnell and Schwab model is that part of the cord has to be left intact — in none of their cases did the regenerating axons pass through the lesion, which would be a prerequisite if the strategy were to be applied to cases in which the entire cord is severed. Likewise, attempts using, for example, pieces of peripheral nerve or fetal spinal cord as tissue bridges for regenerating axons in the completely transected spinal cord, have met with only limited success^{4–6,13}, although most of these experiments have been performed in adult animals.

In the other paper in this issue, Iwashita *et al.*² report some truly remarkable results obtained in neonatal rats. They excised a 1.5–2 mm piece of the lower thoracic cord in 1–2-day-old rat pups and replaced it by a piece of fetal spinal cord of similar length. The pieces were placed in

either the correct or reverse orientation. The successful fetal tissue transplants fused with both transected ends to form what is described as a seamless spinal cord with well-organized white and grey matter, and visible spinal roots. Using antero- and retrograde tracers, Iwashita *et al.* identified extensive axonal connections in both directions across the spinal cord transplants: from the sensorimotor cortex and some functionally essential brainstem nuclei to the lumbar cord, and from the lumbar spinal cord segments to both thalamus and cerebellum. These striking results can in part be explained by the young age of the recipients, and probably also by exquisitely delicate and careful surgery in the pups' tiny spinal cords.

Iwashita *et al.* report extensive and long-lasting functional recovery in the most successful cases; interestingly, such recovery occurred only in animals with fetal spinal cord grafts placed in the correct antero-posterior and dorso-ventral orientation, grafts in reverse orientation being ineffective. Functional analysis was limited to a few selected cases, however, so the results will have to be substantiated by more detailed studies in larger groups of animals.

Nevertheless, the results are intriguing, not least because they can be viewed as a follow-up — long overdue — of some work by Ralph Gerard and collaborators carried out more than half a century ago^{14,15}. Gerard and co-workers described extensive functional recovery after complete spinal-cord transection performed *in utero* or during the early neonatal period. In young rats, recovery was said to be greatly promoted by the insertion of neural or muscle transplants

1. Schnell, L., Schneider, R., Kolbeck, R., Barde, Y.-A. & Schwab, M. E. *Nature* **367**, 170–173 (1994).
2. Iwashita, Y., Kawaguchi, S. & Murata, M. *Nature* **367**, 167–170 (1994).
3. Ramon y Cajal, S. *Degeneration and Regeneration of the Nervous System* (Hafner, New York, 1959).
4. Clemente, C. D. & Windle, W. F. *J. comp. Neurol.* **101**, 691–731 (1954).
5. David, S. & Aguayo, A. J. *Science* **214**, 931–933 (1981).
6. Richardson, P. M., McGuinness, U. M. & Aguayo, A. J. *Brain Res.* **237**, 147–162 (1982).
7. Schnell, L. & Schwab, M. E. *Nature* **343**, 269–272 (1990).
8. Bjerre, B., Björklund, A. & Mobley, W. Z. *Zellforsch.* **146**, 15–43 (1973).
9. Diamond, J., Holmes, M. & Coughlin, M. J. *J. Neurosci.* **12**, 1454–1466 (1992).
10. Tuszynski, M. H., Buzsaki, G. & Gage, F. H. *Neuroscience* **36**, 33–44 (1990).
11. Hagg, T. H., Vahlsing, M., Manthorpe, M. & Varon, S. *J. Neurosci.* **10**, 3087–3092 (1990).
12. Ringstedt, T., Lagercrantz, H. & Persson, H. *Devl Brain Res.* **72**, 119–131 (1993).
13. Reier, P. J., Andersson, F. J., Thompson, F. J. & Stokes, B. T. *J. Neurotrauma* **9**, S223–S248 (1992).
14. Gerard, R. W. & Grinker, R. R. *Arch. Neurol. Psychiat.*, *Chicago* **26**, 469–484 (1931).
15. Sugar, O. & Gerard, R. W. *J. Neurophysiol.* **3**, 1–19 (1940).
16. Gerard, R. W. in *Regeneration in the Central Nervous System* (ed. Windle, W. F.) 229–240 (Thomas, Springfield, Illinois, 1955).
17. Goldberger, M. E., Bregman, B. S., Vierck, C. J. & Brown, M. Exp. Neurol. **107**, 113–117 (1990).
18. Bregman, B. et al. *Exp. Neurol.* **123**, 3–16 (1993).
19. Howland, D. R., Bregman, B. S., Tessier, A. & Goldberger, M. E. *Soc. Neurosci. Abstr.* **17**, 236 (1991).

into the 1–3 mm gap separating the cut surfaces¹⁵.

These results were seriously questioned by later investigators, and the controversy could not be resolved at the time because of the lack of convincing morphological evidence for axonal regeneration across the gap^{4,16}. Then, as now, the main concern was that rats with a severed spinal cord can, unlike humans, regain mobility of their hindquarters. True recovery of supraspinal control of well-coordinated locomotion can thus be revealed only by more detailed analyses; clear criteria for such analyses are now available¹⁷.

Although they demand cautious inter-

pretation, the results of Iwashita *et al.*, as well as work emerging from other laboratories^{18,19}, mean that the time may well be ripe to settle the old argument as to whether extensive functional spinal cord regeneration is indeed feasible under favourable conditions. Or, as Ralph Gerard put it some 40 years ago¹⁶: "The exciting thing, I repeat, is that once one knows regeneration is inherently possible it is just a question of making it work". □

Anders Björklund is in the Section of Neurobiology, Department of Medical Cell Research, University of Lund, S-223 62 Lund, Sweden.

NEUROBIOLOGY

HIV displays its coat of arms

Stuart A. Lipton

ABOUT half of the children and one-third of the adults with acquired immunodeficiency syndrome (AIDS) eventually develop neurological problems, including dysfunction of cognition, movement and sensation, often resulting in frank dementia. The enigma has been that these abnormalities can occur in the virtual absence of neuronal infection by the human immunodeficiency virus type 1 (HIV-1) and in the absence of superimposed malignancy or opportunistic infections. In an elegant study involving transgenic mice, described on page 188 of this issue¹, Toggas *et al.* now place at least one pathway to AIDS neuropathology on firmer ground. They demonstrate that damage in the central nervous system (CNS) can be caused by the HIV-1 coat protein gp120, even in the absence of viral infection.

Eight years ago came the first report of neurological dysfunction during HIV-1 infection², in what was then termed the AIDS dementia complex. Shortly thereafter, it was found that picomolar concentrations of gp120, which could potentially be shed by virus predominantly harboured in CNS macrophages, were toxic *in vitro* to rodent hippocampal neurons³. This work was extended by evidence that gp120 could indirectly trigger a dramatic and potentially lethal rise in neuronal $[Ca^{2+}]$ by releasing toxic factors from activated macrophages/microglia and possibly astrocytes^{4,5}. Further, it emerged that this increase in neuronal $[Ca^{2+}]$ is mediated by gp120 in conjunction with glutamate, acting at *N*-methyl-D-aspartate (NMDA) receptors⁶.

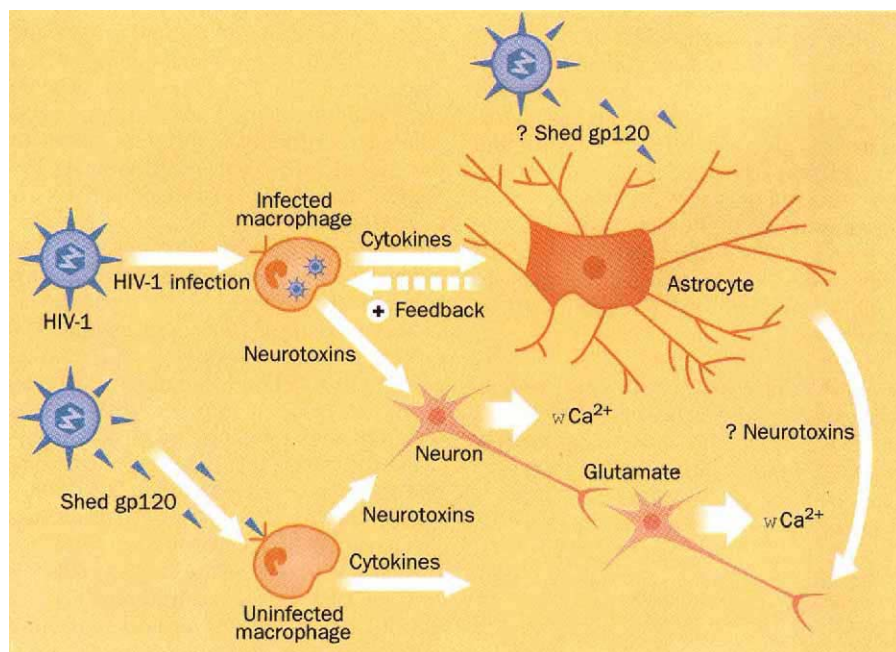
As in a variety of neurodegenerative diseases⁷, excessive intracellular Ca^{2+} is thought to contribute to a final common pathway of neurotoxic events leading to free-radical formation, cellular necrosis and possibly apoptosis; these events in-

clude overactivation of the neuronal enzymes protein kinase C, Ca^{2+} /calmodulin-dependent protein kinase II, phospholipases, proteases, protein phosphatases, xanthine oxidase, endonucleases and nitric oxide synthase. Nitric oxide, in conjunction with superoxide anion, is believed to be involved in one of the toxic pathways activated by NMDA receptor stimulation⁸, and inhibitors of nitric oxide synthase indeed ameliorate gp120-induced neuronal damage *in vitro*⁹.

But doubts then arose about the possible role of gp120 in AIDS dementia. For

example, although gp120 can bind to CD4 on human macrophages it has been argued that this is not true for rodents. Thus, the effects of gp120 in the rodent nervous system might imply the existence of another, as yet unknown, receptor for the coat protein. In addition, other HIV proteins, such as tat, were shown to be toxic in the rodent CNS, raising questions about the specificity of the findings with gp120. In humans, the neurological consequences of AIDS are often accompanied by injury to or loss of neurons, as well as an increase in the number of reactive astrocytes and activated macrophages/microglia.

Toggas *et al.*¹ present encouraging evidence that the neuropathological findings are similar in gp120 transgenic mice. Nevertheless, as the authors themselves acknowledge, this mouse model has its shortcomings. First, the transgene for gp120 is expressed in astrocytes — with the aid of a glial fibrillary acidic protein (GFAP) promoter — rather than in the macrophage/microglial lineage, the cell type predominantly infected in the CNS. Second, it is not possible to demonstrate the existence of gp120 glycoprotein in the brains of AIDS patients or transgenic mice although its messenger RNA is present. This could conceivably be due to the extremely low levels of expression (picomolar or subpicomolar) of gp120 that can cause neuronal injury. In support of this possibility, Toggas *et al.* were able to provide evidence by immunoprecipitation for translation and secretion of transgene-



Current model of HIV-related neuronal injury. Macrophages/microglia infected by HIV-1 or activated by gp120 secrete potential neurotoxins (see text for candidate substances) and cytokines. Neurons are thereby excited, intracellular Ca^{2+} rises and the excitatory neurotransmitter glutamate is released, leading to further neuronal injury in neighbouring neuronal cells. Cytokines released from macrophages contribute to astrogliosis, and astrocytes may in turn send a feedback signal (dashed line) to increase macrophage secretions. Astrocytes may also be directly influenced by gp120 and release substances contributing to neurotoxicity.

derived gp120 protein by astroglial cells in culture, but even then they had to use a transfected C6 glioma cell line to ensure high expression.

In the final analysis, gp120 should be viewed as just one piece of a complicated network of factors and cell types that appear to contribute to neuronal damage in AIDS patients. The main way HIV-1 enters monocytoid cells is by gp120 binding, and so it is not surprising that gp120 (or a fragment of it) can activate uninfected macrophages to release toxic factors similar to those secreted in response to frank HIV infection. Arachidonic acid, platelet-activating factor, free radicals (NO^* , $\text{O}_2^{\cdot-}$), glutamate, quinolinic acid, cysteine, cytokines (tumour necrosis factor- α , interleukin- 1β), and as yet unidentified factors emanating from stimulated macrophages and/or reactive astrocytes may all contribute to the neuropathology¹⁰ (see figure).

What now needs to be done is to learn more about these pathways in order to speed up the development of potential therapeutic agents. To date, vasoactive intestinal peptide (or its congener, peptide T)³, calcium-channel antagonists⁴ and NMDA antagonists^{6,11,12} have all been claimed to ameliorate gp120-induced neuronal damage, and clinical trials are in progress. With the advent of the gp120 transgenic mouse, these and other candidate drugs can now be screened more quickly. Although the commonly used anti-retroviral drug, zidovudine, helps to stem cognitive decline in many AIDS patients, this effect levels off after three months of treatment. The newer nucleoside agents, such as dideoxyinosine (didanosine) and dideoxycytidine, do not penetrate the CNS as well as zidovudine, and so their effectiveness for the neurological aspects of AIDS is questionable. One thing is certain, however — as better systemic treatments become available and AIDS patients live longer, the need to develop more effective drugs to combat cognitive dysfunction becomes more urgent. □

Stuart A. Lipton is in the Department of Neurology, Children's Hospital and Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

Dating the young ocean floor

David M. Christie

DETAILED studies of the petrology and evolution of mid-ocean-ridge volcanoes have long been hampered by a lack of direct, accurate ages of lava samples, even when they are well located in space and from heavily sampled areas. Even in those limited areas where geological mapping by observation from submarines or by remotely operated photo/video techniques has allowed reasonable estimates of flow volumes and relative age relationships, the detailed insights into the history of individual volcanic systems that are available through field mapping of volcanoes on land have remained elusive. The development (or refinement) of techniques to achieve routine, accurate numerical ages for young mid-ocean-ridge lavas has been eagerly awaited. On page 157 of this issue, Goldstein and colleagues¹ describe their use of uranium-series disequilibrium dating in the first systematic application of such techniques to the recent evolution of a mid-ocean-ridge segment.

In 1987, scientists interested in the petrology and evolution of mid-ocean-ridge volcanoes met in Chicago with experts in techniques for determining the ages of geological samples². The object was to foster the development of techniques for dating young mid-ocean-ridge basalts, especially those from within the present normal magnetic polarity interval (since 750 thousand years ago), to provide data for a better understanding of the emplacement and evolution of oceanic crust. Key factors are the frequency and volume of eruptions at different places on the global mid-ocean-ridge system, the degree to which eruptive activity is focused at the spreading axis, and the distribution and longevity of individual magma batches (groups of lavas with a common parentage).

There emerged a consensus that the most immediate promise lay in refining two existing radiometric dating methods, primarily through improvements in the techniques and technology of mass spectrometry. These techniques are uranium- and thorium-series disequilibrium dating, and ^{40}Ar – ^{39}Ar dating. In both cases, the principal challenge for mid-ocean-ridge studies was to resolve the very low concentrations of the relevant radionuclides in basaltic lavas. Since then, although disequilibrium dates have been obtained for mid-ocean-ridge basalts from various localities^{3–5}, progress in applying both techniques to new, detailed sample sets has roughly kept pace.

The first ^{40}Ar – ^{39}Ar results from a sampling programme near the East Pacific Rise ($9^\circ 30'$ N to $11^\circ 20'$ N) were presented

at the 1993 Fall Meeting of the American Geophysical Union in San Francisco⁶. They show generally smooth age increases with off-axis distance in the 200–800 kyr range for which this technique appears best suited. Because uranium-series disequilibrium dating can be applied to samples of ages 0–350 kyr, the two techniques can be used together to provide very effective age resolution for the elusive first 750 kyr of crustal growth.

The decay of ^{238}U , ^{235}U and ^{232}Th to ^{206}Pb , ^{207}Pb and ^{208}Pb respectively is characterized by the existence, in each case, of several short-lived intermediate daughter isotopes. Given sufficient time, all isotopes in each decay chain evolve to an equilibrium state in which all their activities are equal. This equilibrium can be easily upset because the daughter elements differ in their geochemical behaviour and are readily fractionated by melting, crystallization and other magmatic processes. Following a fractionation 'event', the time required for each intermediate parent–daughter pair to return to near-equilibrium is about five times its half-life. Conversely, if a parent–daughter pair can be shown to be out of equilibrium, a maximum age of about five half-lives is rigorously demonstrated. More precise ages can be obtained only if an initial value for the ratio in question can be reliably assumed or established, perhaps by means of an internal isochron (for example, using mineral separates from a single sample to obtain collinear isotopic compositions).

As a practical matter, internal isochrons cannot be resolved for most mid-ocean-ridge basalts because the dispersion of the U/Th ratio among the mineral phases is small. Therefore, the assumption is commonly made that the initial ratios for off-axis basalts were the same as those found in present-day basalts from the nearby axis. This assumption is potentially the main uncertainty of the method, and thorough tests of its validity must be a high priority.

Goldstein and co-authors have used both the ^{238}U – ^{230}Th and ^{235}U – ^{231}Pa parent–daughter pairs to obtain ages for a suite of samples, ranging in age from zero to 75 kyr and up to 4 km off-axis, from the Adventure segment of the East Pacific Rise near $9^\circ 30'$ N. This volcanically very active region achieved notoriety with the discovery, slightly further north, of the 'tubeworm barbecue' site, the aftermath of a very recent eruption^{7,8} (see box on following page). Parallel studies of samples from this region⁹ have shown that off-axis eruptions are surprisingly common and that the lavas from these erup-

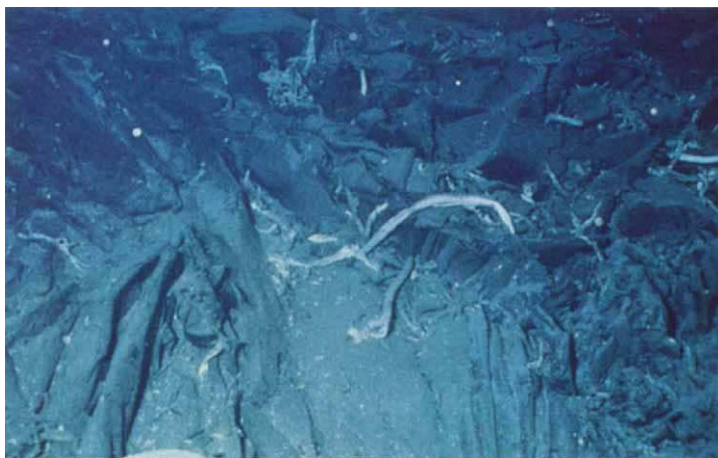
1. Toggas, S. M. *et al.* *Nature* **367**, 188–193 (1994).
2. Navia, B. A., Cho, E. S., Petito, C. K. & Price, R. W. *Ann. Neurol.* **19**, 525–535 (1986).
3. Brenneeman, D. E. *et al.* *Nature* **335**, 639–642 (1988).
4. Dreyer, E. B. *et al.* *Science* **248**, 364–367 (1990).
5. Lipton, S. A. *NeuroReport* **3**, 913–915 (1992).
6. Lipton, S. A., Sucher, N. J., Kaiser, P. K. & Dreyer, E. B. *Neuron* **7**, 111–118 (1991).
7. Lipton, S. A. & Rosenberg, P. A. *New Engl. J. Med.* (in the press).
8. Lipton, S. A. *et al.* *Nature* **364**, 626–632 (1993).
9. Dawson, V. L., Dawson, T. M., Uhl, G. R. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3256–3259 (1993).
10. Epstein, L. G. & Gendelman, H. E. *Ann. Neurol.* **33**, 429–436 (1993).
11. Lipton, S. A. *Neurology* **42**, 1403–1405 (1992).
12. Müller, W. E. G., Schröder, H. C., Ushijima, H., Dapper, J. & Bormann, J. *Eur. J. Pharmac.* **226**, 209–214 (1992).

Aftermath of a sea-floor eruption

VOLCANIC eruptions on the mid-ocean ridges are thought to be the most common on Earth, yet no one has ever seen one in action. Instead we study the aftermath, the lava flows and volcanic cones and ridges built up by these eruptions, trying to deduce what happens when lava bursts forth onto the sea floor. Recent work has moved us much closer to an actual eruption, so that now the immediate consequences of deep-sea volcanism are becoming known. A great difficulty lies in identifying locations along the 75,000 km of ocean ridges where eruptions are occurring this month, or will occur next year. During the past three years, new eruptions have finally been identified on the East Pacific Rise at 9° 50' N and on a segment of the Juan de Fuca Ridge, at 46° 30' N. The first site was discovered serendipitously during dives by the submersible *Alvin*, apparently only days after the eruption. The second eruption was detected in progress by the United States Navy's sensitive Sound Surveillance System hydrophone array, installed during the Cold War to monitor military submarines (*Eos* 74, 619–620; 1993).

The photograph shows a scene from the 'tubeworm barbecue' pit on the East Pacific Rise, 9° 50.6' N, where the volcanic eruption killed a deep-sea hydrothermal vent community (R. M. Haymon *et al.*, *Earth planet. Sci. Lett.* 119, 85–101; 1993). Divers in *Alvin* recovered scorched pieces of tubeworm and other animals in the area. Both the freshness of the tissue and dating of the new lava by ^{210}Po – ^{210}Pb radioactive decay indicated

that this eruption took place only 8–18 days before divers first visited the site on 14 April 1991 (K. H. Rubin and J. D. Macdougall, *Eos* 72, 231; 1991). The ashy sediment was apparently created by a hydrovolcanic explosion that fragmented glassy basalt and a hydrothermal mineral deposit, together with the hapless marine organisms.



In hot water — fresh, corrugated lava and some of its tubeworm victims.

Both eruption sites have now been visited several times by expeditions using remote photographic, mapping and sampling instruments, seismometers and *Alvin* dives. The 9° N flow is a sheet flow extruded onto the floor of the axial summit caldera. It would have been invisible to a 'before-and-after' comparison of sea-floor maps, which have a resolution of only 10 m. The Juan de Fuca eruption, in contrast, created an elongated pillow ridge about 20–25 m high that is clearly identifiable on new bathymetric maps.

Although flowing red-hot lava was not observed, the expeditions did find extraordinarily fresh, glassy lava accompanied by a new type of hydrothermal activity and biology that may be unique to

the initial stages of hydrothermal vent systems (see also R. W. Embley *et al.*, *Geology* 19, 771–775; 1991). At both sites, the new lava flows had voluminous, diffuse hydrothermal fluid venting directly through cracks in the basalt. The temperatures and compositions of the initial fluids at 9° N had unusual and extreme characteristics (high gas content, low salinity, low pH) that pointed to vapour condensation after subsurface boiling. The biology was limited to whitish, flocculant bacterial mats, up to 5 cm thick. The fluid flows disrupted the mats and created a blizzard of white particles suspended in the water for tens of metres. Visits a year later saw a reduction in the diffuse venting and bacterial mats, construction of vent chimneys, and colonization by 'conventional' vent organisms (R. M. Haymon *et al.*, *Eos* 73, 524–526;

1992). An *Alvin* expedition is now under way to check on progress.

Active eruption sites reveal critical processes that occur on short time-scales, particularly the transport of magma from crustal reservoirs to the sea floor, the areal distribution of flows from a single episode, the time between episodes, and the establishment and evolution of hydrothermal circulation and its accompanying biological colonies. These colonies living in eruption zones must stay abreast of the geological changes, as we would like to do.

Jennifer Reynolds

Jennifer Reynolds is at the Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.

tions display a remarkable petrological diversity, including both 'normal' (N-MORB) and 'enriched' (E-MORB) varieties. The new radiometric dates seem to confirm that recent eruptions have occurred as far as 4 km off-axis, a distance equivalent to 60 kyr of sea-floor spreading.

The unexpected abundance and diversity of off-axis volcanism at the AdVenture site is in marked contrast to a second well-studied site near 12° N (ref. 10). At this site, there is little evidence for off-axis volcanism and the spatial distribution of off-axis basalt compositions defines a number of elongated regions that extend for 15–40 km, parallel to the spreading axis. In each such region, basalt composi-

tions are consistent with derivation from a common parent magma composition, and it is believed that they represent the original petrological zoning of the East Pacific Rise axis. If so, samples in each zone should be similar in age and there should be a regular increase in age with distance off-axis. Such predictions should be readily testable by these newly refined dating techniques.

Segment-to-segment contrasts such as

these illustrate the complexity that underlies even the East Pacific Rise, arguably one of the Earth's simpler volcanic systems, and emphasize the continuing need for precise and accurate age determinations. □

David M. Christie is in the College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA.

- Goldstein, S. J., Perfit, M. R., Batiza, R., Fornari, D. J. & Murrell, M. T. *Nature* 367, 157–159 (1994).
- Batiza, R., Duncan, R. A. & Janecky, D. *Dating Young MORB. Workshop Rep.* (JOI Inc., Washington DC, 1988).
- Rubin, K. H. & Macdougall, J. D. *Earth planet. Sci. Lett.* 101, 313–322 (1990).
- Goldstein, S. J., Murrell, M. T. & Williams, R. W. *Earth planet. Sci. Lett.* 115, 151–159 (1993).

- Volpe, A. M. & Goldstein, S. J. *Geochim. cosmochim. Acta* 57, 1233–1241 (1993).
- Duncan, R. A. & Hogan, L. G. *Eos* 74 (suppl.) 687 (1993).
- Haymon, R. M. *et al.* *Eos* 72, 505; 507 (1991).
- Haymon, R. M. *et al.* *Earth planet. Sci. Lett.* 119, 85–101 (1993).
- Perfit, M. R. *et al.* *Geology* (in the press).
- Reynolds, J. R. *et al.* *Nature* 359, 493–499 (1992).

Ancient relationships

Dan Hultmark

INSECTS look nothing like vertebrates, and their organ systems seem to be built on entirely different principles. Nevertheless, as we get a better understanding of how these systems operate at the molecular level, unexpected similarities are emerging. Among them must now be counted similarities in the respective immune defences, as reported in two recent papers^{1,2}.

insect defensins and lysozymes, that accumulate in the blood³. It is the study of this reaction that has led to the discovery of a relationship between insect and vertebrate immune responses.

The first indication of such a connection came from Sun *et al.*⁴. Working with the genes for the inducible antibacterial proteins in the cecropia moth, they noticed that many of these genes share a common

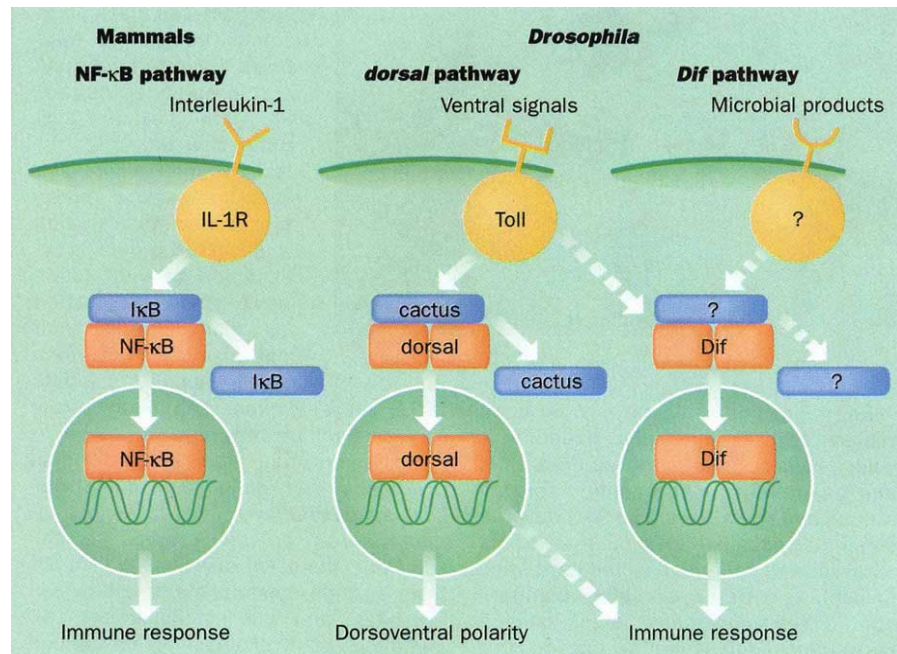
various gene constructs in transgenic flies, and in a *Drosophila* blood-cell line.

One member of the Rel family of transcription factors has previously been described from *Drosophila*. This is the product of the *dorsal* gene, which is involved in the specification of dorsoventral polarity in the embryo⁹. There are in fact several surprising similarities between the signal system used by the fly to instruct the embryo which side is up and which is down, and the one that mediates the effects of interleukin-1 (IL-1) in the mammalian immune system (see figure). In both cases, inactive Rel proteins (NF- κ B or dorsal) are bound to I κ B-like inhibitors (I κ B or cactus) in the cytoplasm. After activation, the Rel proteins are translocated into the nucleus where they mediate the transcriptional activation of target genes. Furthermore, the ventral-specific signal to the embryo is mediated by a membrane receptor, encoded by the *Toll* gene, which is homologous in its cytoplasmic part to the IL-1 receptor (IL-1R).

The findings of Ip *et al.*¹ are the fruits of a collaboration between workers in the fields of *Drosophila* embryology and *Drosophila* immunology. A homologue of *dorsal* was discovered that is inactive in the embryo but becomes expressed during later stages of development. This gene, dubbed *Dif* (dorsal-related immunity factor), is preferentially expressed in the fat body, the main site of synthesis of the antibacterial proteins, and it was therefore considered to be a good candidate for the immunoresponsive transcription factor. Ip *et al.* now demonstrate that Dif protein binds specifically to the κ B-like element from the cecropin gene, whereas the dorsal protein does not. Like other Rel proteins, Dif is normally cytoplasmic, but becomes nuclear in the fat body after induction with bacteria. Amazingly, some cross-talk may be possible with components of the dorsoventral pathways; a mutant in which Toll is constitutively activated induces nuclear localization of Dif, as well as increased expression of the *Dif* gene.

At the same time, Reichhart *et al.*² found that the *dorsal* gene itself may also be involved in the immune response. This gene is expressed at a low level in fat body cells of larvae and adults, but levels of messenger RNA increase when the immune system is stimulated by lipopolysaccharide. Like Dif, the dorsal protein becomes concentrated in the nucleus upon induction. Furthermore, dorsal binds to the κ B-like element in the dipterin gene *in vitro*; and, when expressed in transfected cells, dorsal protein can stimulate transcription from reporter gene constructs that include this element.

Several questions remain. Although dorsal protein is able to bind to the dipterin promoter, and to stimulate



Comparison of three pathways of gene activation mediated by Rel proteins. The receptor and the inhibitor that interact with Dif are hypothetical. NF- κ B is a heterodimer of two Rel proteins, p50 and p65, and other active hetero- and homodimers may form from these and other Rel proteins. It is not known whether dorsal and Dif also form similar complexes.

A hallmark of the immune system in vertebrates is the acquired immunity based on the clonal selection of T and B cells. Such systems have not been found in insects, and it has even been argued that the speed of insect reproduction would make an immune system unnecessary. But microorganisms often have doubling times of an hour or less, allowing them to grow to saturation overnight, whereas the generation times of insects are typically measured in months. In vertebrates, acquired immunity takes days or weeks to develop. So both insect and vertebrate must depend on more rapid defences, such as the poorly understood mechanisms of 'natural', or innate, immunity.

Insects respond promptly to invasion by bacteria, which are recognized and attacked by phagocytic blood cells. Within hours a humoral response is activated, with the generation of a diverse set of broad-spectrum antibacterial proteins and peptides, such as cecropins, attacins,

upstream motif, similar to the binding site for NF- κ B, a mammalian member of the Rel family of transcription factors. In mammals, NF- κ B and other Rel proteins have a central role in the transcriptional activation of immune-related factors such as immunoglobulins, interleukins and the proteins of the acute phase response⁵. The same transcription factors are also exploited by the human immunodeficiency virus when it infiltrates the human immune system.

Sun and Faye⁶ went on to show that known inducers of the insect antibacterial response, such as bacterial lipopolysaccharide, activate a nuclear factor that binds to the conserved κ B-like motif, and that this factor cross-reacts with antibodies against NF- κ B. Next, taking advantage of *Drosophila* as an experimental system, two groups demonstrated the functional importance of the κ B-like motif in the *Drosophila* cecropin⁷ and dipterin genes⁸, by introducing

transcription, this is not a prerequisite for induction because dipterin is still inducible in *dorsal* mutants. Obviously, there must be some redundancy in the system. Furthermore, *dorsal* does not bind to the *cecropin* promoter. But the mammalian Rel proteins are believed to form various homo- or heterodimers with slightly different sequence specificities; so Dif and *dorsal* may conceivably form similar complexes, with each other or with other factors in the *Drosophila* cells, to form the transcription factors that activate the immune response *in vivo*.

Why a similar (perhaps even the same) system is used for immunity and for the formation of the dorsoventral body axis remains a mystery. Ip *et al.* argue that the induction of innate immune responses is the original function of the Rel proteins. One may speculate that when insects evolved eggshells as an adaptation for a terrestrial lifestyle, the immune system may have been recruited for the com-

munication between the embryo and its immediate surroundings, to ensure that the main axes of the embryo are properly aligned with those of the eggshell.

The discovery of this ancient immune system in *Drosophila* could be of much wider significance. It may give us a key to unravel the mechanisms and evolutionary roots of our own innate immunity. □

Dan Hultmark is in the Department of Molecular Biology, Stockholm University, S-106 91 Stockholm, Sweden.

1. Ip, Y. T. *et al.* *Cell* **75**, 753–763 (1993).
2. Reichhart, J.-M. *et al.* *C. R. Acad. Sci. Paris, Sciences de la vie/Life Sci.* **316**, 1218–1224 (1993).
3. Hultmark, D. *Trends Genet.* **9**, 178–183 (1993).
4. Sun, S.-C., Lindström, I., Lee, J.-Y. & Faye, I. *Eur. J. Biochem.* **196**, 247–254 (1991).
5. Liou, H. C. & Baltimore, D. *Curr. Opin. Cell Biol.* **5**, 477–487 (1993).
6. Sun, S.-C. & Faye, I. *Comp. Biochem. Physiol.* **B103**, 225–233 (1992).
7. Engström, Y. *et al.* *J. molec. Biol.* **232**, 327–333 (1993).
8. Kappler, C. *et al.* *EMBO J.* **12**, 1561–1568 (1993).
9. StJohnston, D. & Nüsslein-Volhard, C. *Cell* **68**, 201–219 (1992).

SUPERCONDUCTIVITY

Progress down below

Z. Fisk

The past month has witnessed a flurry of reports hinting at superconductivity near room temperature (see last week's News and Views by Grant¹). Yet on page 146 of this issue we find Cava and colleagues² reporting the observation of superconductivity in an intermetallic material with $T_c = 23$ K. One might well wonder why anyone bothers to report such a modest T_c . But this concoction, identified here only as part of a polyphase material containing yttrium, palladium, boron and carbon, includes an intermetallic whose superconducting transition temperature essentially ties with the highest T_c found for any other intermetallic compound, that of thin-film Nb_3Ge . This record T_c was achieved in 1974. We now have the first progress in superconducting intermetallics in 20 years.

Among pure elements, the highest T_c is 12 K (ref. 3), in lanthanum under pressure. In binaries, it is possible to double this. Why is it not possible to achieve further increases in ternary or higher intermetallics? It probably is, but nobody, except perhaps Bob Cava's group and a few others, has any clue how to make an efficient search of the enormous ternary and higher phase space for superconductors. The method of the pioneers in this area, Hulm, Matthias and Geballe, was much like Thomas Edison's method of trying everything, occasionally focused by some empirical trend. And indeed, although Cava's group was undertaking a systematic search, this new superconducting composition was one mistakenly made

with the wrong stoichiometry. It is refreshing that a poor idea of where to look is often more useful than a good one in this search for higher T_c s: it is more open to the unexpected.

One thing we do know is that certain intermetallic structures are much more favourable than others for superconductivity. The A15 structure of Nb_3Ge contains a large number of compounds with respectable T_c s. What Cava *et al.* now give us, or so we hope, is a new structure to play with chemically. They do not know what this structure is, or even if it is a ternary or quaternary (although developments to be published next week by the same group^{4,5} suggest the latter). Carbon often stabilizes intermetallic structures, for example the Cu_3Au structure, which has a stoichiometry close to the Y-to-Pd ratio here.

There seems little reason to doubt that Cava *et al.* do have a superconductor. Intermetallics are not fraught with the traps for the unwary found in the oxides. But the superconducting phase could be filamentary: a surprisingly small amount of superconducting material in grain boundaries can give zero resistance and large shielding. Measurements made on a crushed sample, combined with a specific heat measurement through T_c , will go a long way towards addressing this.

In follow-on work, Cava *et al.*⁴ and Siegrist *et al.*⁵ have explored other quaternary borocarbides, and find superconductivity near 17 K in $LuNi_2B_2C$. The crystal structure is tetragonal and closely related

to that of $ThCr_2Si_2$, whose family of member compounds reaches biblical proportions. This new borocarbide contains carbon occupying a crystallographic site that is empty in $ThCr_2Si_2$, one of the useful ways in which to create new derived structures. Another way in which this is done is by partially substituting a different element for one that occupies two crystallographically inequivalent sites in the structure. In fact, this is how the $ThCr_2Si_2$ structure is derived from the $BaAl_4$ structure. At the very least, we can now expect extensive attempts to add carbon to the many known and interesting compounds with the $ThCr_2Si_2$ structure, as well as detailed investigation of the low-temperature magnetic properties of the other rare-earth analogues.

Siegrist *et al.*⁵ observe that their 17 K superconducting borocarbide is part of a homologous series of compounds of general formula $(LuC)_m(NiB)_n$, where the $m=n=1$ member is not superconducting, and the $m=1$, $n=2$ member is. This series consists of layers and invites comparison to the layering in the cuprates, which also can exist in homologous series. Another interesting similarity to the cuprates is in the temperature dependence of the electrical resistivity: it is common to see so-called electrical resistivity saturation near room temperature in the higher- T_c intermetallics, but neither the cuprates nor $LuNi_2B_2C$ seem to show this, perhaps a reason to think about alternatives to the electron-phonon mechanism here.

Why should we care about this new range of materials? There are at present serious problems in using cuprates for high-current applications. The large structural anisotropies are a metallurgical nightmare for conductor forming, and at present, weak links and/or low flux pinning restrict the useful operating range of cuprates for such purposes to temperatures far below their T_c s. Anisotropy is not likely to be a problem with intermetallics, and although intermetallics tend to be brittle, their precursors are often fairly easily formed into useful shapes. A new class of intermetallic holds the potential for even higher T_c s and higher upper critical fields than are available in the commercial superconducting wires now in use, Nb–Ti and Nb_3Sn . Increases of 50 per cent or more do not seem out of the question, given recent advances in other non-cuprate superconductors such as the fullerenes. We are reminded that the $T_c=11$ K of Art Sleight's perovskite $Ba(Pb_{1-x}Bi_x)O_3$ was more than doubled⁶ following Mattheiss's suggestion⁷ of sub-

1. Grant, P. M. *Nature* **367**, 16 (1994).
2. Cava, R. J. *et al.* *Nature* **367**, 146–148 (1994).
3. Maple, M. B. *et al.* *Phys. Rev. Lett.* **23**, 1375 (1969).
4. Cava, R. J. *et al.* *Nature* (in the press).
5. Siegrist, T. *et al.* *Nature* (in the press).
6. Cava, R. J. *et al.* *Nature* **332**, 814–816 (1988).
7. Mattheiss, L. F., Gyorgy, E. M. & Johnson, D. W. *Phys. Rev.* **B37**, 3745–3746 (1988).

stituting the barium site partially with potassium, one of the very few cases I am aware of where a superconductivity theorist made a useful prediction for increasing T_c . And such increases would have real application. We may find in the ternary and quaternary materials ways to thwart the structural and other instabilities which so often appear to limit the T_c s achievable in intermetallics, instabilities which rather surprisingly do not seem to limit the T_c s of the cuprates.

The investigation of the properties of

bulk intermetallic phases has been rather neglected in condensed-matter physics and solid-state chemistry in recent years. This discovery by Cava *et al.* serves to remind us that there is a vast and quite unknown area in intermetallic materials which will hold many surprises. □

Z. Fisk is in the Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA, and Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

AGRICULTURE

Crops and climate change

John Reilly

WILL global climate change over the next century seriously limit agricultural production and cause widespread famine? Rosenzweig and Parry, writing on page 133 of this issue¹, investigate this possibility and find that although the location of food production may shift, global food prospects do not worsen significantly. There seems to be no 'disaster threshold' for global food production.

This finding is at odds with the language of the Framework Convention on Climate Change, which calls for the world to curb emissions and avoid dangerous greenhouse gas concentrations. The notion of a threshold implies that there is some level of trace gases where the effects are so devastating that we must avoid it at whatever cost, but that we can edge up to this limit as long as we do not exceed it. With the emphasis instead on adjustment and relocation of agricultural production, the debate on climate-change policy shifts to manageable rates of change and to concerns about the responsibility of the world community towards those areas that suffer significant losses.

Impacts

Rosenzweig and Parry's work looks at the agricultural effects to be expected from temperature and precipitation changes associated with a doubling of atmospheric carbon dioxide concentration (or equivalent trace gases), together with the effects of direct 'CO₂ fertilization'. They investigate three different general circulation models (GCMs), each of which gives a change of several degrees in mean global temperature in the next century, and go on to look at crop yields, possible adaptations and a model of world food trade.

Theirs is the first extensive effort to produce a consistent estimate of the impacts of climate change worldwide from specific GCM scenarios. Earlier work tested sensitivities and relied on disparate estimates of the impact on crop yields

with, for many areas of the world, very limited information². The basis of this study is similar, however, in that it relies on detailed crop growth models for specific geographical points. These few points form the basis for extrapolating to states or provinces, nations and regions. The suite of crop response models used by Rosenzweig and Parry was limited to a few (important) grain crops, the results being extended to other crops with reference to the literature. An alternative approach would be to use databases of global geographical information, such as global climate and soils, to estimate crop potential on the basis of the specific climate in each grid and estimates of how that climate will change³.

Adaptation was introduced according to the judgement of local agricultural scientists in each participating country. As such, adaptation response was only as creative and complete as the imagination of the agronomist concerned. Any small group of individuals studying the situation today is unlikely to foresee the full range of possible adaptations — crop breeding, technical advances and more — that might be developed over a hundred years of gradually changing climate. Further, agronomists in the team generally experimented with adaptations at only the most affected site in their country (or region of responsibility). As a simplification, yield losses were halved for the nation if adaptation was partial, and set to zero if it was complete, so that adaptation never outweighed losses. But if adaptation measures could completely or nearly completely offset losses at the most severely affected sites, the same adaptation might well lead to yield increases at the less severely affected sites.

Rosenzweig and Parry's view that farmers will only try to maintain yields against losses but not adapt to increase them is probably over-pessimistic. The Green Revolution and the long-term trend of productivity growth (roughly 2.1

per cent per annum) in US agriculture both testify that farmers will try to increase yields if it pays them to do so. A fuller consideration of adaptation may show that, globally, several of the GCM scenarios could give overall increases in production.

Moreover, the focus on crops that are staples of temperate agriculture may have led to an overestimate of the negative impacts in tropical regions. Thus, their conclusion that developing countries will be more severely affected, although compelling, remains open to further investigation using methods that can consider beans, root crops, sugar cane, and various fruit and vegetable crops better suited to warm climates.

Finally, the economic impact of agricultural changes may strongly depend on whether a country is an agricultural exporter or importer^{4,5}. Agricultural exporting countries fared well during the 1970s when various world commodity prices were high. If climate change leads to high agricultural prices, then developing countries such as Thailand or Argentina, which are major agricultural producers and exporters, could do well, whereas major consuming areas, regardless of the direct impact of climate on agriculture in the region, could suffer. Tempting as it is to simplify the world into 'developed' and 'developing' nations, this may give a misleading picture of the impact on individual countries.

Hunger

A unique contribution of Rosenzweig and Parry's study is the attempt to link the risk of hunger specifically with climate projections. It is generally recognized that current food production capacity is adequate to avoid famine and malnutrition, but famine occurs because available food does not always get to those most in need. Wars and political upheaval are the primary cause of current famines (as in Somalia and Bosnia). In other words, famine and food shortage is a short-term, unexpected (although recurring) phenomenon, whereas climate change is a long-term trend, at least as described by the current generation of GCMs. This trend may, indeed almost certainly will, exacerbate famine potential in some areas. As Rosenzweig and Parry demonstrate, many local production changes may create a global picture of sufficient food, but regional populations continue to be very vulnerable to famine.

This agricultural study contains the core of a debate that will play out in the current Intergovernmental Panel on Climate Change and future negotiations on the Framework Convention. Certainly, it seems callous to balance the human catastrophe of famine, with thousands of deaths and severe malnutrition, against an economic cost of dollars per tonne of

carbon. Shouldn't the humanitarian spirit be willing to spend the necessary money to limit emissions and avoid these potential famines? But from our current experience we know that current famines are a choice, not an inevitability built of too little food.

Given the possibility (but not assurance) of redistributions of resources to mitigate the locally severe effects of climate change, rational cost-benefit calculations cut little ice in the international negotiations on climate change. Those who see their prospects for famine or coastal destruction increasing will find little comfort in the assurance that the increase in trace gases is consistent with a balancing of costs and benefits. At the same time, the world as a whole can ill afford an extremely costly effort to control emissions if the effects are locally severe but reasonably cheap, in principle, to compensate. Such funds might save far more lives if they were spent on AIDS research, eliminating water-borne diseases or improving childhood nutrition.

For a reasoned climate change policy to be agreed, those at particular risk must have some assurance that their unavowed losses will be compensated. This is not a straightforward exercise. It will be difficult to separate the drought or hurricane attributable to climate change from the drought or hurricane that would have occurred anyway. It seems that for the world to deal with the problem of climate change rationally, it must be prepared to provide a basic level of security to all its citizens.

The long horizon of climate change, with the potential for famine to increase, suggests that now is the time (if one needed another reason) to make a concerted effort to eliminate famine and malnutrition. Poverty, subsistence agriculture, poorly developed social support networks and weak governmental processes appear to be their preconditions; training, improved health delivery, education, economic development and improved agriculture production techniques are their solution⁶. With success the future can be one without famine or malnutrition, regardless of whether and how the climate changes. □

John Reilly is at the Economic Research Service, 1301 New York Avenue NW, Washington DC 20005, USA.

And they all look just the same

Paul Calvert

SOME solution precipitations give particles with a very narrow range of sizes, rather loosely called monodisperse. There is a standard explanation for this uniform size which dates from work by LaMer in the 1950s¹. A series of papers has now made it clear that LaMer's explanation is wrong, but leaves only an enigma in its place.

There has been a flurry of interest in these powders, starting about five years ago when it was thought that they would be good for making high-strength ceramics. At present, this route would be too expensive; but given a cheap method of producing the particles in abundance, applications in medical diagnostics, composites and paints would quickly follow. Early papers all referred to the LaMer model, where reaction in solution progressively generates a solute which finally exceeds the concentration needed for nucleation. Once this occurs, a number of particles start to grow, rapidly reducing the concentration of solute to a level below that needed for further nucleation.

As a result, the theory goes, all particles start at the same time and grow to the same size. Egon Matijevic of Clarkson College has been publishing on monodisperse inorganic particles for many years, with many beautiful micrographs of identical spheres, cubes, rods and ellipsoids. In a review in April 1993, Matijevic surprised many of us by saying that the LaMer model was wrong and discussing a number of alternatives².

The most studied system for formation of monodisperse particles was described by Stöber in 1968 (ref. 3). He hydrolysed a dilute solution of tetraethoxysilane at high pH in alcohol to form particles of a silica gel. The particles were spheres of almost identical size, with a diameter which varied from about 0.1 µm to 1 µm depending on the pH and concentrations used.

For some years, a dissenting voice to the LaMer model has been that of Zukoski and co-workers at the University of Illinois⁴, who say that nanometre-sized colloidal particles can be seen in solution both before and after nucleation in the

Stable relationships

It's doubtful that members of pony clubs usually take much interest in biological research, but results described by Claudia Feh and Jeanne de Mazières in the December issue of *Animal Behaviour* (46, 1191–1194; 1993) will provide them with compelling reading. Feh and de Mazières have made behavioural and physiological observations on Camargue horses, such as those pictured here, with the aim of finding

out more about the reasons for mutual grooming. It could, for instance, in certain circumstances or species be to remove parasites or in some way to reduce tension between individuals. By observing a herd of horses during the early summer (when ectoparasite prevalence is at its lowest) the authors identified a highly preferred target for grooming; this is the base of the neck, as demonstrated by these two subjects. Grooming takes the form of scratching the partner's skin with the upper incisors, and Feh and de Mazières found that over half such 'grooming contacts' were directed at this one small area which occupies barely one-hundredth of a horse's body surface. They went on to measure heart rate in eight adults and eight foals when natural grooming was imitated by scratching the same spot by hand, and found that it was significantly reduced (by some 12 per cent, on average) against that observed in horses before grooming or horses that were groomed in a less favoured place. So what could the physiological basis of this calming effect be? Feh and de Mazières point out that a large ganglion, which is linked to certain efferent cardiac pathways, lies close to what might be called the C-spot, but a direct connection remains to be established.

T.L.

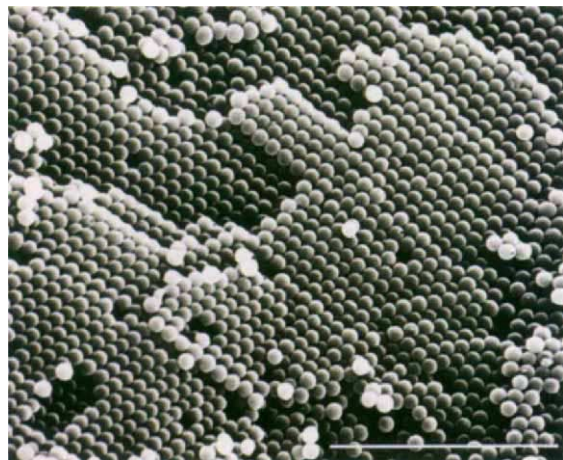
Only Horses

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

1. Rosenzweig, C. & Parry, M. L. *Nature* **367**, 133–138 (1994).
2. Kane, S., Reilly, J. & Tobey, J. *Clim. Change* **21**, 17–35 (1992).
3. Leemans, R. & Solomon, A. M. *Clim. Res.* **3**, 79–96 (1993).
4. Reilly, J., Hohmann, N. & Kane, S. *Glob. envir. Change* (in the press).
5. Reilly, J. & Hohmann, N. *Am. Econ. Rev.* **83**, 306–312 (1993).
6. Ruttan, V. *Sustainable Growth in Agriculture: Poetry, Policy and Science*, Staff Pap. P91-47 (Department of Agricultural and Applied Economics, Univ. Minnesota, St Paul, 1991; revised 1993).

Stöber system. These primary particles are essentially the result of polymerization to form cross-linked silica gel particles. The 0.5- μm spheres are the result of aggregation of the primary particles rather than growth by molecular addition. Zukoski *et al.* produce a model that predicts narrow particle size distributions as a result of this aggregation.

Although electron microscopy does show primary particles, there is the possibility of these being artefacts of the drying process. Cryoelectron microscopy shows that the primary particles in the Stöber



Submicrometre silica spheres formed by the Stöber process. Scale bar, 10 μm . (Courtesy of M. Jiang, Fudan University.)

system are not dense but loose polymeric structures⁵. A similar titania-forming system does however show small dense particles.

In support of aggregation, electron microscopy also shows a fine-scale substructure within many of the varied types of particles. For instance, rod-shaped particles can be seen to be made up of little rods. Zukoski *et al.* point out that this ordering of nanometre-sized particles is similar to the crystallization of globular proteins. Also disturbing, from the viewpoint of the LaMer model, is Matijevic's remark that spherical particles can be found with strong crystalline X-ray patterns, although this could be explained by crystallization subsequent to the particle formation. He also observes that nucleation can continue after the initial stages of the reaction, which is contrary to LaMer's model.

van Baaderen, van Geest and Vrij from the University of Utrecht⁶ disagree. They measured the growth of Stöber silica on seed particles with a bimodal size distribution, derived from two separate precipitations. By following the change in the ratio of the two radii, they showed that the radial growth rate is independent of radius; this, they argue, is consistent with growth being limited by the rate of hydrolysis of tetraethoxysilane and not with aggregation. An induction time for precipitation is observed because time is needed to build up a sufficient concentra-

tion of hydrolysed species before condensation occurs on the surface of the seed. Also, addition of dissolved electrolytes would be expected to enhance an aggregation process by reducing the range of coulombic repulsions between charged particles, but no such effect is seen.

A new paper swings the pendulum back to aggregation. Briois *et al.*⁷ have used small-angle X-ray scattering and EXAFS (extended X-ray absorption fine structure) to study the precipitation of cerium oxysulphate from heated cerium sulphate solutions, a system that others have used

to make monodisperse particles. During the heating stage, they see particles form and stabilize as 1.8-nm monodisperse spheres. Several hours later, this suspension starts to precipitate larger particles. The fact that a monodisperse colloid forms, stabilizes and then starts to grow again, suggests that the governing process is one of aggregation rather than reaction. As the paper describing this second stage is in preparation, we must wait for the other shoe to drop.

The charm of the LaMer model is that the source of

the narrow size distribution is quite clear. In the aggregation models there is no such simple clarity in the source of monodispersity. The other standard system that produces monodisperse particles is emulsion polymerization, which is the source of the latex particles that are used to calibrate electron microscopes. A burst of free radicals initiates polymer chains which grow until all the monomer is consumed. As each chain grows within a surfactant micelle, they are kept separate and each particle should be a single polymer chain. There have been attempts to draw parallels with Stöber silica, but they are not yet convincing.

Different systems seem to show quite different behaviour. There should be a single explanation for why monodisperse particles are formed but there is evidence against each available model. For the moment this is a case of the more you know, the less you understand. □

Paul Calvert is in the Department of Materials Science and Engineering, University of Arizona, Tucson, Arizona 85712, USA.

1. LaMer, V. K. *Ind. Engng Chem.* **44**, 1270–1277 (1952).
2. Matijevic, E. *Chem. Mater.* **5**, 412–426 (1993).
3. Stöber, W., Fink, A. & Bohn, E. J. *Colloid interf. Sci.* **26**, 62 (1968).
4. Zukoski, C. F., Chow, M. K., Bogush, G. H. & Look, J.-L. *Mat. Res. Soc. Symp.* **180**, 131–140 (1990).
5. Bailey, J. K. & Mecartney, M. L. *Mat. Res. Soc. Symp.* **180**, 153–158 (1990).
6. van Baaderen, A., van Geest, J. & Vrij, J. J. *Colloid interf. Sci.* **154**, 481–501 (1992).
7. Briois, V. *et al. J. Mat. Sci.* **28**, 5019–5031 (1993).

Flowing free

SHIPS ploughing through the ocean, hydraulic pumps, oil-lubricated bearings — all have to overcome boundary-layer friction. The liquid next to a solid surface sticks on that surface; liquid further away can only move by viscous shear against this boundary layer. Daedalus now has a way out.

He points out that the spherical C_{60} molecules of buckminsterfullerene rotate freely in their crystal lattice. It was once thought that the resulting molecular ball-bearing surface should be utterly frictionless. Not so: solid surfaces are far too rough for such tiny ball bearings to run on them. But a liquid surface is much more hopeful. Imagine, for example, a ship coated with buckminsterfullerene. The freely rotating molecules will be spun up by the drag of the water flowing past them. They cannot quite spin freely, for their rotation is quantized. But their range of accessible rotation rates includes peripheral velocities of the order of a few metres per second: just right for marine transport. So a fullerene-coated ship should slip through the water on molecular ball bearings. It would forge frictionlessly ahead, while its micro-spinning solid surface would in effect remain stationary with respect to the water. Viscous drag would vanish.

This elegant idea is annoyingly sabotaged by the slow oxidation of fullerenes in air. Daedalus argues that partially hydrogenated or fluorinated derivatives should be more stable, and would still show free rotation. Indeed, chemically modified fullerenes suggest an even more cunning scheme. Suppose a spherical fulleroid molecule was mounted on some sort of 'axle', so that it could spin only in one plane. The ideal molecule would be a linear polyfullerene whose spherical moieties were joined together like free-spinning beads on a string. Such a long-chain polymer could easily be molecularly aligned by spreading it over a flat surface. The coated surface would then have a unique 'easy direction' for fluid flow — the free-spinning direction of the fullerene moieties. The coating would enforce linear streamline flow in one defined direction. Flow in any other direction, including random turbulence, would be damped by viscous drag.

In paint form, the new polymer will gladden the hearts of ship designers, chemical engineers and everyone else concerned with fluid flow. Properly applied, it will eliminate the twin curses of viscous drag and turbulence from pumps, pipes, ships and bearings. Vast amounts of energy, now wasted uselessly in stirring and shearing liquids, will be saved.

David Jones

Early vertebrate colour vision

SIR — A feature of retinæ in lizards, turtles and birds is the intercalation of a coloured oil droplet between the inner and outer segments of cone photoreceptor cells^{1,2}. The colour in the droplets absorbs particular wavelengths of light, thereby narrowing the spectral sensitivity of the four opsins (visual-pigment proteins) that are located in the outer segments. The result is a tetrachromatic visual system with a broad sensitivity that extends from near-ultraviolet (~350 nm) to infrared (~750 nm) (birds^{3,4}, turtles^{5,6}). The Australian lungfish, *Neoceratodus forsteri*, belongs to the order *Dipnoi* which diverged from the main vertebrate stock in the early Devonian. *N. forsteri* fossils have

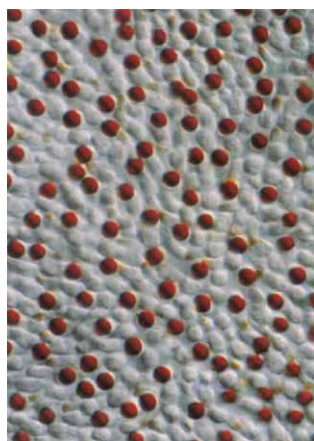


FIG. 1 Photomicrograph showing the array of coloured oil droplets in the cone photoreceptors of an unstained *Neoceratodus* retina. The small colourless droplets are not evident in this plane of focus. Lungfish were deeply anaesthetised, their eyes quickly removed, and the retinæ immersed for 24 hours in 4% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4). The retinæ were then flatmounted with the photoreceptor layer uppermost, and coverslipped in 50% glycerol in 0.1 M phosphate buffer. Magnification, $\times 150$.

been dated at well over 100 million years, making this the oldest extant vertebrate species⁷. Here I show that *N. forsteri* has coloured droplets in its cone photoreceptors, suggesting that the common ancestors of lungfish and land vertebrates also had coloured oil droplets and, perhaps, tetrachromatic vision.

In the eight *Neoceratodus* retinæ that I have studied, cones are distributed at densities of 2,500–4,600 mm⁻² and they compose 50–56% of the photoreceptor population, the remainder being rods: 65–75% of cones possess red oil droplets, 5–10% have small colourless droplets, and the rest (15–25%) have a golden granular pigment at the base of the outer segment, instead of an oil droplet (Fig. 1). In some *Neoceratodus* the variety of oil droplets increases towards the retinal

edge to include orange. The density of red droplets is always reduced in regions where orange droplets are present, but there is no reduction in the density of small colourless droplets or cones containing golden granular pigment.

The cone oil droplets in *Neoceratodus* have a larger diameter (6–15 μ m) than those in turtles (3–10 μ m)^{8,9} or birds (1–6 μ m)^{9,10}. The droplet colours observed in *Neoceratodus*, as well as colourless droplets and cones with a yellow granular pigment in the ellipsoid, are found in birds^{3,4,9,10} and turtles^{5,6,8,9}, but not all colours are present in every species.

Walls¹ proposed that coloured oil droplets appeared very early in vertebrate evolution, and that when species adopted a nocturnal habit, colour was lost from their droplets, or the droplets were eliminated. Once lost, the colours could not be regained, even if the descendants of a nocturnal animal colonized a diurnal niche, as has occurred among some mammals and frogs. This proposal was not widely accepted because of the restriction of coloured oil droplets to a small part of the vertebrate series. The restricted distribution could be explained more simply if coloured oil droplets originated about 250 million years ago in the common ancestors of turtles, birds and reptiles (Fig. 2). The discovery of coloured oil droplets in the lungfish provides much stronger support for the notion that coloured oil droplets were present in the lobe-finned ancestors of land vertebrates (*Sarcopterygii*) more than 400 million years ago (Fig. 2).

The number of cone opsins possessed by *Neoceratodus* is unknown. However, in all vertebrates that have been studied, colour vision is achieved through the use of two or more different photopigments, rather than by coloured oil droplets in combination with a single opsin^{2,4}. The fact that coloured oil droplets are restricted to those vertebrates which have four or more cone opsins⁴ implies that the colour in the droplets only conveys an

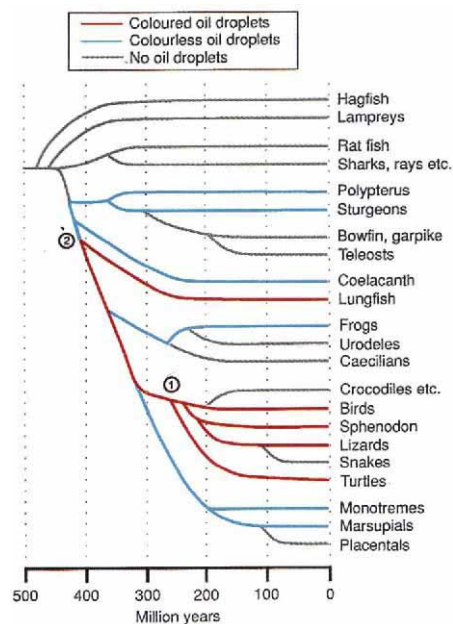


FIG. 2 Approximate origins of the main extant vertebrate taxa, based on data in refs 1, 2, 11. Taxons lacking oil droplets in their cone photoreceptors are indicated in grey, those with colourless droplets in blue, and those with coloured droplets in red. Until now, the distribution of coloured droplets was consistent with an origin about 250 million years ago. The finding of coloured droplets in lungfish supports a common origin 400 million years ago.

advantage when many opsins are present. This perspective favours the idea that *Neoceratodus* has four cone opsins, and raises the possibility that the ancestral *Sarcopterygii* had tetrachromatic vision.

Stephen R. Robinson

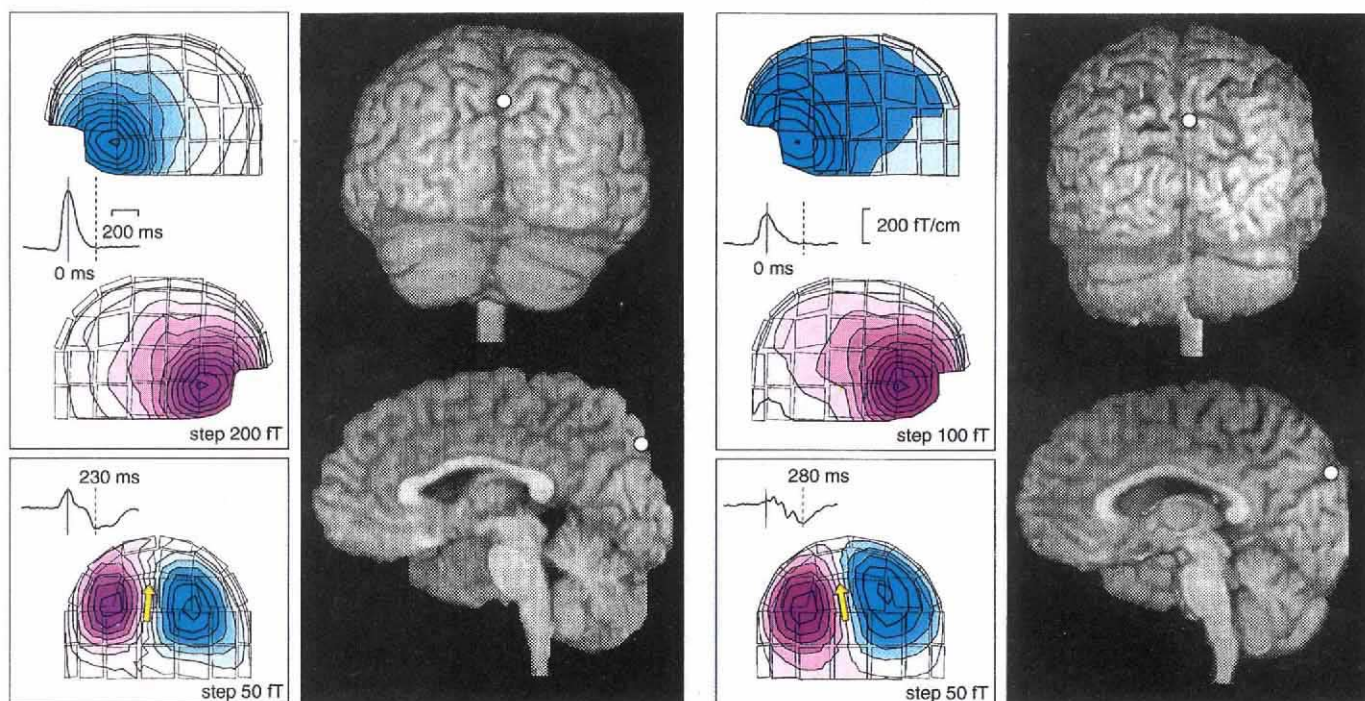
Vision, Touch and Hearing Research Centre,
University of Queensland,
Brisbane,
Australia 4072

Visual stability during eyeblinks

SIR — During each eyeblink we lose sight of the visual world for more than a tenth of a second without usually perceiving the discontinuity. A suppression of visual sensitivity during blinks^{1,2} explains why darkening is not seen but it is not sufficient to account for the continuity of visual perception. Here we report on activation of the human parietal lobe after voluntary blinks. The posterior parietal cortex — an essential part of the spatial working memory system³ — may have a role in maintaining a stable image despite the interruption of visual input during blinks.

We studied five adult subjects (ages 27–34 yr; 3 females) with a helmet-shaped 122-channel SQUID (superconducting quantum interference device) magnetometer (Neuromag-122™). The subjects

1. Walls, G. L. *The Vertebrate Eye and its Adaptive Radiation* (Bloomfield Hills, Michigan, 1942).
2. Crescitelli, F. in *Photochemistry of Vision* (ed. Dartnall, H. J. A.) 245–363 (Springer, Berlin, 1972).
3. Partridge, J. C. J. *comp. Physiol.* **A165**, 415–426 (1989).
4. Bowmaker, J. K. in *The Perception of Colour: Vision and Visual Dysfunction Vol. 6* (ed. Gouras, P.) 108–127 (Macmillan, Basingstoke, 1991).
5. Lipetz, L. E. in *The Visual System* (eds Fein, A. & Levine, J. S.) 107–132 (Liss, New York, 1985).
6. Ohnaka, T. J. *comp. Neurol.* **237**, 145–154 (1985).
7. Kemp, A. in *Vertebrate Palaeontology of Australasia* (eds Vickers-Rich, P. et al.) 465–496 (Pioneer Design Studio, Melbourne, 1991).
8. Kolb, H. & Jones, J. J. *comp. Neurol.* **209**, 331–338 (1982).
9. Kawata, A. et al. *Photochem. Photobiol.* **56**, 1157–1166 (1992).
10. Csernti, P., Szel, A. & Rohlich, P. *Invest. ophthalm. Vis. Sci.* **30**, 74–81 (1989).
11. Marshall, C. & Schultze, H.-P. *J. molec. Evol.* **35**, 93–101 (1992).



Field maps and MRI surface renderings for two subjects. The magnetic field patterns are shown during the maximum blink signal (upper boxes) and during the peak of the posterior response 230 and 280 ms later (lower boxes). The helmet-shaped sensor array is viewed from three angles (left, right and back). The patterns are based on simultaneous recordings with 122 sensors; each of the sensor units (squares in the figure) houses two orthogonal planar gradiometers. Red, magnetic flux out of the head; blue, flux into the head; the separation between isocontours ('step') is indicated in each box.

Examples are shown of averaged magnetic signals from field extrema. The arrows show the sites and orientations of the equivalent current dipoles for the posterior signal patterns; dipole strengths are 23 and 26 nA m⁻¹ for subjects 1 and 2, respectively. The white dots superimposed on the three-dimensional MRI surface rendering of the same subjects' brains (viewed from the back and from the mesial surface of the right hemisphere) illustrate the equivalent current dipoles for the late responses. The current flow is perpendicular to the course of the parieto-occipital sulcus.

fixated on a cross or a small picture and blinked voluntarily once every 3–5 s. Altogether 3×5 min of data were gathered, with two 1-min pauses in-between. The magnetic signals were averaged off-line using the onset slope of the vertical electro-oculogram as the trigger.

Strong magnetic signals were seen close to both orbits during the blinks (see figure): the magnetic flux emerged from the scalp over the right side and entered the head on the left side. More than 200 ms later (220–285 ms in the five subjects), when blink signals were no longer observed, new responses emerged over the upper posterior part of the head. Similar late responses have been observed previously in scalp EEG records⁴. The posterior field patterns were satisfactorily accounted for by current dipoles, which corresponded in strength to sources of strong evoked responses⁵.

Superposition of the source locations on the individual three-dimensional magnetic resonance imaging (MRI) reconstructions (made for two subjects; see figure) implied activity close to the parieto-occipital sulcus, probably in the posterior association cortex (area 7). Conventional early visual evoked responses occurred at a much lower site. No corresponding magnetic signals were detected when

blinking occurred in complete darkness (studied in one subject), indicating that an interruption of a visual stimulus is required for the appearance of this response. The relevant activity was obviously in the posterior parietal cortex, which receives afferent input from the striate cortex through a dorsal visual stream, a pathway considered to be involved in spatial rather than object vision^{6,7}. No corresponding activation of the ventral stream was observed.

The posterior parietal cortex is reciprocally connected with prefrontal cortical areas⁸ which seem to underlie spatial working memory³. It has been suggested that the monkey posterior parietal cortex updates continuously information about the nature and structure of visual objects in the ego-centred space^{9,10}. The present results are consistent with similar pathways and mechanisms in man, inferring that the posterior parietal lobe is kept informed also about eyeblinks. Most blinks occur unnoticed by the subject, and blink-related information thus seems to be processed unconsciously.

We hypothesize that the observed blink-related responses in the human posterior parietal cortex are related to spatial working memory, necessary for maintaining a continuous image of the

environment despite the 0.1-s loss of visual input during each blink. Such continuity is essential for stable visual perceptions. The latency of the parietal activation suggests that the eyeblink is reacted to after its occurrence, not in advance in connection with the motor command.

R. Hari, R. Salmelin, S. O. Tisari
Low Temperature Laboratory,
Helsinki University of Technology,
02150 Espoo, Finland

M. Kajola
Neuromag Ltd., c/o Low Temperature
Laboratory, 02150 Espoo, Finland

V. Virsu
Department of Psychology,
PO Box 11, 00014,
University of Helsinki,
Finland

1. Volkman, F., Riggs, L. & Moore, R. *Science* **207**, 900–902 (1980).
2. Ridder III, W. H. & Tomlinson, A. *Vis. Res.* **33**, 1795–1802 (1993).
3. Wilson, F. A. W., O'Scalaidhe, S. P. & Goldman-Rakic, P. C. *Science* **260**, 1955–1958 (1993).
4. Berg, P. & Davies, M. *Electroenceph. Clin. Neurophysiol.* **69**, 1–5 (1988).
5. Hämäläinen, M. *et al. Rev. Mod. Phys.* **41**, 413 (1993).
6. Mishkin, M., Ungerleider, L. & Macko, K. *Trends Neurosci.* **6**, 414–417 (1983).
7. Horwitz, B. *et al. J. cog. Neurosci.* **4**, 311–322 (1992).
8. Cavada, C. & Goldman-Rakic, P. S. *J. comp. Neurol.* **287**, 422–425 (1989).
9. Goodale, M. A. & Milner, A. D. *Trends Neurosci.* **15**, 20–25 (1992).
10. Stein, J. F. *Behav. Brain Sci.* **15**, 691–700 (1992).

Optimal design of honeycombs

SIR — In 1964, L. F. Toth delivered an entertaining lecture entitled "What the bees know and what they do not know"¹. In it, he analysed the geometry of the bees' honeycomb in terms of the efficient use of wax and hence the minimization of the total area of the surfaces involved. When viewed in such terms, the architecture used by the bees is not quite

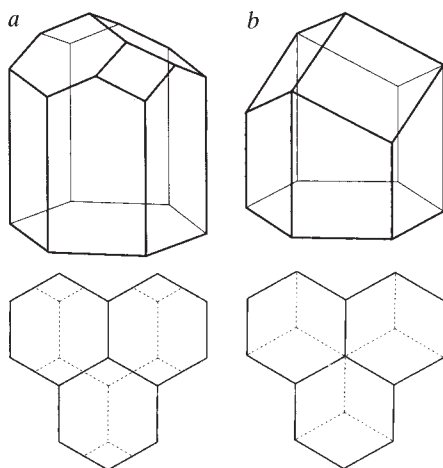


FIG. 1 Alternative structures for the honeycomb interface. *a*, Toth structure; *b*, beehive structure.

optimal, but the possible improvement on the bees' design is very small.

Toth went on to describe a superior design for the honeycomb interface, while noting that there was no proof that his choice was in any sense optimal in the biological context. It is closely related to Kelvin's proposed optimal division of space into periodic cells².

We have undertaken an experiment which demonstrates Toth's proposed structure, invites a generalization of his problem to foam structures of finite liquid content, and shows that the bees' choice is reinstated as superior whenever the liquid fraction is sufficiently high. The structures in question are illustrated in Fig. 1.

We can explore the analogous problem for liquid foam by introducing equal-sized bubbles (here with approximately 2-mm diameter) of a detergent solution between two glass plates, the plate separation being such that a double layer of bubbles is trapped. Each layer forms an array of hexagonal cells. A foam of low liquid content consists of thin films whose junctions are readily idealized as lines. In this case, which corresponds to Toth's geometrical picture, we observe the structure proposed by him (Fig. 2).

If we wet the foam by the addition of more liquid, the junctions between films are thickened to form what are called Plateau borders. The conditions under which surface energy is to be minimized are accordingly changed. What happens

as the foam is progressively wetted is quite dramatic. At a certain point the Toth structure becomes unstable and there is a sudden switch to the configuration favoured by the bees (Fig. 2). Such a switch also takes place in the reverse direction, as liquid is removed.

This structural transition relates to current ideas of foam structure in three dimensions^{3,4}. It will provide a particularly convenient test of quantitative theories of topological changes due to instabilities, and will be further analysed by us in that context.

D. Weaire

R. Phelan

*Department of Pure and Applied Physics,
Trinity College, Dublin, Ireland*

1. Toth, L. F. *Bull. Am. math. Soc.* **70**, 469–481 (1964).
2. Thomson, W. (Lord Kelvin) *Phil. Mag.* **24**, 503–514 (1887).
3. Weaire, D. *Phil. Mag. Lett.* (in the press).
4. Weaire, D. & Phelan, R. *Phil. Mag. Lett.* (in the press).

How populations persist

SIR — How much suitable habitat or resources does a population require to persist or, to put it another way, what is the eradication threshold of the population? According to one approach first developed in epidemiology¹, the simplest estimate of the eradication threshold of a disease (specifically, how large a fraction of the population, failing to be covered by a vaccination programme, would allow the disease to persist) is the observed fraction susceptible to the disease at equilibrium. In more general terms, an estimate of the eradication threshold for a population is simply the unused amount of its limiting resource (susceptible individuals being the limiting resource for the multiplication of a disease organism).

Lande^{2–4} derived an estimate of the eradication threshold for territorial species, that is, an estimate of the proportion of a region which must consist of suitable territories if the population is to persist, and applied this estimate to the northern spotted owl, *Strix occidentalis caurina*, which has been the subject of intense interest by conservationists. Lande's estimate of the threshold is the fraction of sites which are currently unoccupied although they are suitable (see ref. 5 for a more detailed discussion). Given that suitable breeding territories were taken to be the limiting resource for territorial species, this is the epidemiological estimate. The fact that the many biological differences between owls and say, weasles, do not seem to affect the eradication threshold estimate prompts the investigation of a rather different ecological model to probe its generality further.

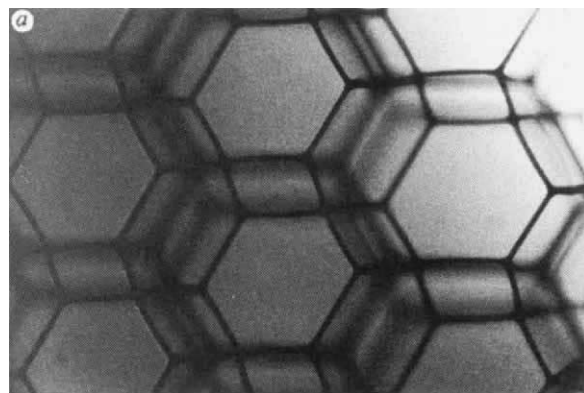
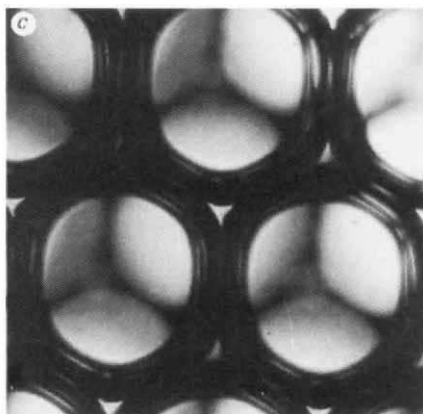
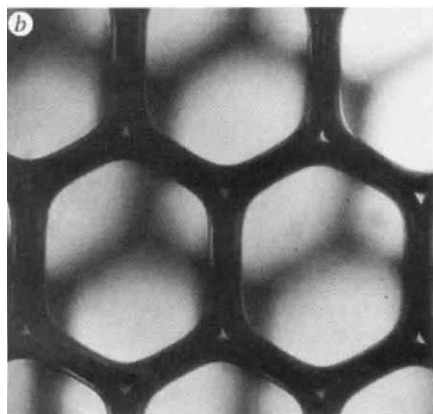


FIG. 2 *a*, Interface of Toth structure in foam viewed at a slight angle. *b*, View of surface before and *c*, after transition to the alternative (beehive) structure.



Consider a predator–prey relationship described by the following model:

$$\begin{aligned}\frac{dv}{dt} &= vr(1-v/K) - \alpha vp \\ \frac{dp}{dt} &= \alpha \beta vp - \mu p\end{aligned}\quad (1)$$

v and p are the densities of the prey and predator, respectively, K is the carrying capacity of the prey, μ is the death rate of the predator, r is the intrinsic growth rate of the prey and α and β are parameters controlling the capture of prey and their conversion into predator biomass. A common objection on this sort of model is that it is too simple to be of any use in the 'real world'; but in fact, the model consists mostly of irrelevant detail from the point of view of calculating an eradication threshold.

The eradication threshold of the predator is determined by the minimum prey carrying capacity, K , that can sustain a predator population (the amount of habitat that would support K prey in the absence of predation), which is straightforward to calculate. Setting the time derivatives in equations (1) equal to zero and solving for v^* and p^* , the equilibrium abundances of prey and predator,

$$\begin{aligned}v^* &= \mu/\alpha\beta \\ p^* &= \frac{r(1-\mu/\alpha\beta K)}{\alpha}\end{aligned}\quad (2)$$

The eradication threshold is that value of K such that $p^*=0$, that is, $K=\mu/\alpha\beta=v^*$, the equilibrium abundance of the prey. So the eradication threshold is simply the unused amount of the limiting resource.

To reveal what accounts for this result, I shall rederive the estimate for the predator–prey case as derived in ref. 1 in the epidemiological context. The derivation in ref. 1 consists solely of a verbal argument which can be applied to many biological models with an obvious change of wording. Equilibrium in the system is achieved when the density of prey is reduced to that level at which a predator gives rise to a single progeny before dying. This is the equilibrium density v^* . In the absence of predation, the density of the prey would be K . So, speaking loosely, an amount $K-v^*$ prey is 'used up' in keeping the predator population alive. Biologically, this cannot be a negative number. Hence, v^* immediately presents itself as the minimum value of K required to

sustain a predator population. None of the species-specific details of predator hunting efficiency, death rates, efficiency of conversion of prey biomass into predator biomass, and so on, enters into the estimate of the eradication threshold. As is demonstrably the case in epidemiological contexts, details that at first glance seem important in fact cancel each other out (for review of theory and data, see ref. 1). In addition to showing what is not important, this simple approach also makes it easier to see what features are important for estimating eradication thresholds^{1,5}.

It would be trite to observe that epidemiology and conservation biology are both motivated by the same concern — the eradication of species — were it not for the fact that there are many areas of theoretical overlap between the two disciplines. Some other areas are described in ref. 6.

Sean Nee
AFRC Unit of Ecology
and Behaviour
Department of Zoology,
South Parks Rd,
Oxford OX1 3PS, UK

Viral burden in HIV infection

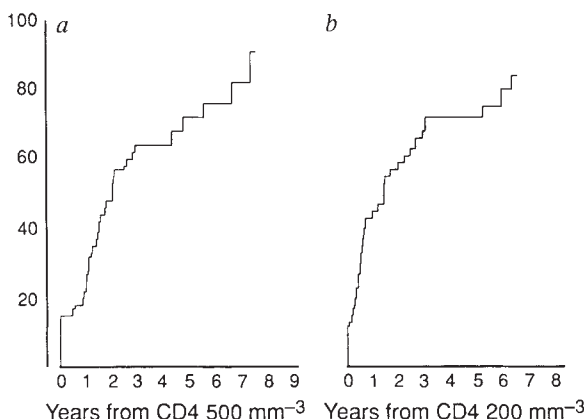
SIR — McLean and Michie¹ and Garry and Fermin² argue in Scientific Correspondence that the direct cytopathic effects of the virus could well be the main cause of

whether virus levels are measured in plasma or in cells (or, based on the little evidence available, lymphoid tissue³) and whether using quantitative PCR (polymerase chain reaction) or endpoint dilution culture^{4,5}. In sharp contrast, the rate of CD4 lymphocyte loss tends, on average, to be no more rapid late in the infection than it is earlier (see figure). This inconsistency cannot apparently be explained by differences in viral pathogenicity. The *in vitro* properties of virus isolated late in infection suggest that it is unlikely to be less able to kill CD4 cells *in vivo* than virus present earlier in the infection⁶.

As well as having fundamental implications for understanding HIV pathogenesis, the weakness of the association between viral load and the rate of peripheral blood CD4 lymphocyte count decline gives some indication of the potential effect of antiviral

drugs. Therapies that induce substantial (about 10–100-fold) reductions in viral load might be expected to produce only small reductions in the rate of CD4 lymphocyte decline. Consistent with this prediction are data from placebo-controlled clinical trials of Zidovudine (AZT), which show an initial increase in the absolute number of CD4 cells following therapy but similar rates of decline in the Zidovudine and placebo arms thereafter⁷.

Andrew N. Phillips, Caroline A. Sablin, Jonathan Elford, Margarita Bofill*, Vince Emery†, Paul D. Griffiths†, George Janossy* & Christine A. Lee‡
Departments of Public Health,
Immunology*, Virology† and
the Haemophilia Centre‡,
Royal Free Hospital School of Medicine,
London NW3 2PF, UK



Kaplan Meier estimates of the percent of haemophilic individuals having a CD4 lymphocyte count below 250 mm⁻³ (a) or 100 mm⁻³ (b) according to the number of years after the count has first fallen below 500 mm⁻³ or 200 mm⁻³ respectively. The time taken to decline from a count of 500 mm⁻³ to a count of 250 mm⁻³ (a fall of 250 mm⁻³) is similar to the time taken to decline from a count of 200 mm⁻³ to a count of 100 mm⁻³ (a fall of 100 mm⁻³), thus illustrating that the rate of CD4 lymphocyte count decline is, if anything, less rapid late in HIV infection than earlier. Study methods have been described previously⁸.

CD4 lymphocyte count decline during HIV infection. If such is the case, however, why does the rate of CD4 lymphocyte loss not increase concomitantly with the enormous increase in viral burden observed during the course of HIV infection? The viral burden late in HIV infection (for example when the CD4 lymphocyte count is around 200 mm⁻³) has consistently been found to be, on average, 10–100 times greater than it is earlier (for example when the CD4 lymphocyte count is around 500 mm⁻³). This is the case

1. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford University Press, 1991).
2. Lande, R. *Science* **241**, 1455–1460 (1988).
3. Lande, R. *Am. Nat.* **130**, 624–635 (1987).
4. Lande, R. *Oecologia* **75**, 601–607 (1988).
5. Lawson, J. H., Nee, S., Letcher, A. & Harvey, P. H. in *Large Scale Ecology and Conservation* (eds Edwards, P. J., May, R. M. & Webb, N.) (Blackwell, Oxford, in the press).
6. Nee, S., Holmes, E. C., May, R. M. & Harvey, P. H. *Phil. Trans. R. Soc. B* (in the press).

1. McLean, A. & Michie, C. *Nature* **365**, 301 (1993).
2. Garry, R. F. & Fermin, C. D. *Nature* **365**, 301–302 (1993).
3. Pantaleo, G. *et al.* *Nature* **362**, 355–358 (1993).
4. Piatk, M. *et al.* *Science* **259**, 1749–1754 (1993).
5. Bieniasz, P. D. *et al.* *AIDS* **7**, 803–806 (1993).
6. Asjo, B. *et al.* *Lancet* **ii** 660–662 (1986).
7. Cooper, D. A. *et al.* *N. Engl. J. Med.* **329**, 297–303 (1993).
8. Phillips, A. N. *et al.* *Lancet* **337**, 389–392 (1991).

Book case history

Alexander G. Bearn

Companion Encyclopaedia of the History of Medicine. Edited by W. F. Bynum and Roy Porter. Routledge: 1993. Pp. Volume 1, 778; Volume 2, 997. £150, \$199.95.

THE dangers of editing an encyclopaedia, even a small one, are formidable. It took Diderot 20 years and many volumes to complete his *Encyclopédie* and his coeditor D'Alembert gave up in despair 15 years before its publication in 1772. The *Encyclopaedia Britannica*, devised by a "society of gentlemen in Scotland", was completed in three volumes in 1771 and is still with us; the latest edition in 1992 running to 32 volumes.

W. F. Bynum and Roy Porter call these two volumes a "companion encyclopaedia". Avowedly neither comprehensive nor encyclopaedic, their work fits neatly between the genre of the typical 'encyclopaedia' and that of 'collected essays'. Its encyclopaedic dimension consists rather in the number of essays collected than in the comprehensiveness of its coverage. The *Companion Encyclopaedia of the History of Medicine* is a fascinating miscellany of 72 essays of varying length written by 36 carefully selected scholars, mostly British or American, in the history of medicine.

In an important introductory essay entitled "The Art of Science and Medicine", the editors have interpreted their subject in its evolving historical, social and cultural contexts. It is this broad sweep of medical history that sets this encyclopaedia apart from others in the field.

It is divided into seven somewhat arbitrary parts varying from the ambitiously titled Theories of Life, Health and Disease to more narrowly focused chapters on clinical medicine. The section on body systems includes succinct historical essays on basic disciplines such as anatomy, physiology, biochemistry and immunology that are important scientific platforms on which medicine necessarily rests.

An essay on the doctor-patient relationship opens the section on clinical medicine and contains one chapter in which Robert C. Olby considers the constitutionality of disease in enlightening and judicious prose. Christopher C. Booth provides an admirable essay on the role of clinical research in understanding the nature of human disease.

Essays on Arab Islamic medicine, Chinese medicine and Indian medicine are a particularly welcome and illuminating corrective to the more usual emphasis on Western culture.

In one of the more difficult essays, Arthur Caplan struggles with the distinction between health and disease. Although such a distinction may be a

pragmatic necessity, particularly in the social context of governed allocation of limited resources, the concept is biologically unsatisfying. The view frequently expressed by René Dubos, particularly in his book *The Mirage of Health*, that complete freedom from disease is almost incompatible with the process of living, seems more rewarding.

The scholarly essay on psychotherapy by Sander Gilman outlines its chequered history. The ranks of those who cling to orthodox Freudian psychotherapy have

entrenched in the Indian subcontinent, South-East Asia and China, which comprise half the world's population. Almost a century has passed since Patrick Manson identified the anopheles mosquito as the transmitting agent for malaria. Malaria still kills 3,000 children a day in Africa. Airport malaria, particularly among baggage handlers, is well documented and improvements in airline schedules have enabled mosquitoes to become world travellers. Diseases previously regarded as geographically confined have become global. Only smallpox appears to have been eradicated.

The diversity of subjects not ordinarily considered in books of this nature — hospitals, women and medicine, nursing, architecture, literature and medicine, medical philanthropy, to give only a few examples — make this not only a valuable and lively work of reference, but a book to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Aspects of medical history — *The Surgeon* by Jan Sanders van Hemessen (c.1504–66).

steadily shrunk and the view expressed by Peter Medawar more than 20 years ago that "psychoanalysis will remain for ever one of the saddest and strangest landmarks in the history of twentieth-century thought" has more adherents.

The history of contagion and infectious diseases, discussed in more than one essay, is traced with erudition through the centuries. The overall decline in mortality from infectious disease over the past hundred years has been impressive. In his chapter on demography, R. M. Smith challenges T. McKeown's view that this decline has been almost exclusively due to improvement in the environment on the grounds that he largely ignored demographic variations in mortality.

Microbes, old and new, remain formidable agents of disease. Tuberculosis has returned and been made more deadly by strains resistant to all known antibiotics. It is estimated that AIDS, unknown 15 years ago, will affect 40 million people by the end of the decade and is already firmly

be kept close at hand and read with pleasure as well as enlightenment. For those who wish to extend their knowledge, the editors and authors have provided well chosen selected references. This encyclopaedia is an invaluable and scholarly source and deserves to be widely read by those interested in the broader aspects of medical history.

"All sciences", the historian C. V. Wedgwood (in her collected essays, *History and Hope*, 1987) reminds us, "are devoted to the quest for truth; truth can be neither apprehended nor communicated without art. History therefore is an art, like all the other sciences." This well indexed encyclopaedia, edited by two of the leading English scholars in the field, will inform and stimulate all those interested in the art and science of medical history. □

Alexander G. Bearn is at The Rockefeller University, 1230 New York Avenue, New York 10021-6394, USA.

How to press the point

Roland Pease

Hitting the Headlines: a Practical Guide to the Media. By Stephen White, Peter Evans, Chris Mihill and Maryon Tysoe. *British Psychological Society*: 1993. Pp. 144. £8.99.

WHEN the British public woke up at the end of last year to find that the Universe had been taken over by billions of black holes, it was not some malignant extraterrestrial that had been at work. Rather, it was the power of the press release, revealed in glorious Technicolor.

Mike Hawkins's interesting but unexceptional paper on dark matter in *Nature* (366, 242–245; 1993) would have drawn little attention had not the British Science and Engineering Research Council, just weeks before its institutional demise, put out a sensational invitation to the science press corps to learn the “answer to cosmology's biggest mystery”, suggesting “large numbers of black holes” could be “the dominant component of the Universe”.

For science writers, the story was a dream — ‘sexy’, straightforward in concept and with an easy-to-grasp bottom line. Mike Hawkins's willingness to say in public what others would rather keep between consenting professionals in the privacy of the coffee room was a key to the story's popularity. But for many scientists the episode epitomized all that is dangerous in tangling with the media — the tendency to sensationalism, superficiality and irresponsibility.

Finding the fine line between getting publicity for science and exposing it to ridicule is what *Hitting the Headlines* is all about. The first author, Stephen White, is director for information for the British Psychological Society, which has a long record of helpful and stimulating stories for the British media. The other three are professional science journalists. Together, they have assembled a nuts-and-bolts picture of the British press and broadcasting media, how to deal with them, and how to use them. Most of the lessons — on preparing press releases, handling interviews, writing articles — apply anywhere in the world.

If there's a moral to be drawn from the book, it's encapsulated in two rules the authors put forward: “the media knows exactly what it wants from you”; and “you have to put yourself in their hands”. The point is that science journalists know who they are writing for, or broadcasting to, and at what level to pitch the story. Their very purpose is to report the latest (and best) in science. And they hate inaccuracy

and irresponsibility as much as anyone: it does them no good with their peers or their sources.

If they have a hard time, it's because for every Mike Hawkins who might be accused of overselling his results, there are many more scientists who will crush a good story by being unable to speak plainly to the press. Scientists, used to caution and qualification when speaking to their colleagues, find it hard to drop the habit when dealing with the broader public; jargon becomes such second nature they don't realize they're using it. (Even a simple term such as amplification can cause problems, showing how cultures can be divided by a common language. Any

biologist will talk happily about amplifying genes, for example. But most people will be quite perplexed by the concept — can you really make genes louder?)

The common complaint is that there's not enough science reported in the media. The authors of *Hitting the Headlines* argue that coverage is really very reasonable, given the vast range of human activity for the papers and broadcasters to choose from. Nonetheless, they have also prepared an easy and readable prescription for improving matters. Over to you. . . □

Roland Pease is with the Science Unit, BBC World Service, Bush House, London WC2B 4PH, UK.

Getting into deep water

Martin Angel

Global Marine Biological Diversity: A Strategy for Building Conservation into Decision Making. Edited by Elliott A. Norse. *Island Press*: 1993. Pp. 383. \$50 (hbk), \$27.50 (pbk); UK distributors: *Earthscan*, £24.95 (pbk).

MAINTENANCE of diversity and sustainable development have become the two central tenets of world conservation. Both appeared as major objectives in the first World Conservation Strategy, but there was a third in the original strategy, the

species of their marine counterparts, the copepods. The differences in functioning become accentuated once the depth becomes too great for the seabed to influence processes within the euphotic zone, so that the dominant factors influencing the ecosystems are hydrodynamic.

This book, sadly like all the other strategic documents on biodiversity, almost completely overlooks the 51 per cent of the Earth's surface that is covered by water 3,000–6,000 metres deep. There is no mention of ocean trenches. Hydrothermal vents are mentioned because of

their strange and prolific communities, based on chemosynthetic production. But the species-richness of these communities is low relative to the surrounding barren regions. More new species of nematode worms and foraminiferans have been found in a couple of cubic centimetres of abyssal sediment from the North Atlantic than from all the vents investigated so far. There is no mention that each week the world's vents emit more heavy metals than industries throw away globally; does this suggest possible alternative waste disposal policies?

On the evidence of this book the conservation movement seems too divorced from the scientific base from which must come the knowledge it seeks, the technology it requires for monitoring and the ability it needs to understand change and to predict ecosystem response. This is a strategy for a static world, a world in which any change is bad and any change is the result of human activities. Present community structure depends on its history and resilience to change, but how

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hot spot — hydrothermal vent community.

conservation of ecological process, which has become increasingly overlooked. Yet the importance of the roles played by abiotic and biotic ocean processes in global geochemical budgets probably makes this third tenet the most relevant to ocean conservation. Compared with terrestrial ecosystems, oceanic ecosystems are very different in their structure and function. For example, there are more than a million species of insects known from the terrestrial realm, but fewer than 2,000

Peter Ryan/Scripta/Science Photo Library

diversity is related to function is poorly understood. Marine communities have important roles in the carbon cycle, in regulating atmospheric greenhouse gases such as methane or dimethyl sulphide and in breaking down and detoxifying organic and inorganic materials. Is a policy of concentrating conservation effort on species-rich hotspots the correct one for trying to keep the globe working? For example, the most serious short-term threat to oceanic ecosystems, the effects of increased ultraviolet-B radiation resulting from the ozone hole, has been glossed over.

There is a recognition of the need for more taxonomists to be trained, but to which groups should priority be given? The development of networks is considered the best way to disseminate new knowledge, but there is no recognition of the power of new technologies. Taxonomy is haunted by its past. Descriptions of species appear in abstruse journals. Type material is seldom readily available and much has disintegrated over the years. Identification keys are often so complex and poorly illustrated that only experts can work them. These data need to be organized, simplified and made readily available, a task being undertaken by the Expert Centre for Taxonomic Identification (ETI) at the University of Amsterdam using optical disk (CD-ROM) technology.

Perhaps more puzzling is that there is no mention of GOOS (Global Ocean Observing System), which is intended to observe the oceans using satellites, robotics, moorings and traditional surface vessels to provide a forecasting system similar to that provided by meteorologists. Because oceanic systems have 'memories' of at least 2 years, compared with about a week for the atmosphere, the prospects for accurate long-term predictions of weather and climate are much more realistic. Developing a global strategy must involve the effective deployment of limited resources necessitating a predictive capability.

This is a useful and informative contribution towards the development of a global marine strategy for biological diversity. It has too many shortcomings to be a definitive statement. It is a committee document: the input by 100 authors is acknowledged together with reviews by representatives of 40 nations. Even so, it has a strong North American bias. Nevertheless, it is an important first step along a long and complex road towards redressing what Robert May has described as the terrestrial chauvinism of the conservation movement. □

Martin V. Angel is at the Institute of Oceanographic Science, Deacon Laboratory, Wormley, Godalming, Surrey GU8 5UB, UK.

Physics without apparatus

Daniel J. Kevles

A Mind Always in Motion: The Autobiography of Emilio Segrè. By Emilio Segrè. University of California Press: 1993. Pp. 332. \$37.50, £25.00.

A MIND Always in Motion is Emilio Segrè's account — published four years after his death in 1989 — of his personal life and his life in physics. He wrote the book as something of an anticipated consolation for his family, while justifying it on grounds of his engagement at a high level in the grand adventure of twentieth-century science. It is absorbing, moving in places and frequently revealing. Segrè noted in his preface, "I have not sought to display manners and tact I never had, and I have tried to treat myself no better than any one else." He ably succeeded in these purposes.

Segrè was certainly close to the centre of the middle third of this century's adventure in physics. Having grown up in Tivoli and Rome, the product of a prosperous Jewish family, he was an apprentice in physics with the Fermi group and, in 1938, at the age of 33, moved to the United States, another gift to American physics

from Hitler's Europe. He spent most of the rest of his life in Berkeley, California, first as a member of Ernest O. Lawrence's radiation laboratory, then as a member of the University of California faculty. An expert in radiochemistry, he was deeply involved early in the Second World War in the discovery and isolation of plutonium, and he further contributed to the development of the atomic bomb as a member of the Los Alamos laboratory. In 1959 he shared the Nobel prize in physics for having conclusively demonstrated the existence of the antiproton in an experiment done in 1955 at the Berkeley radiation laboratory using the bevatron.

Segrè candidly relates the obstacles and hazards he encountered along the road to his ultimate success. Although counting himself fortunate to be part of Fermi's group — he was supported by an ancient sinecure that paid him formally as Conservator of the Tuning Fork — he recognized the truth in the remark of his father, a shrewd observer of men, that "you live off Fermi's crumbs". His appointment to a professorship in Palermo, in 1935, made him feel liberated from his subordinate status in the Rome institute, but the



NOT coconut shy — the giant coconut crab, from the Fijian island of Vatuvara, climbs palms to cut off coconuts with its powerful claws. *Nomads of the Wind: a Natural History of Polynesia* by Peter Crawford accompanies a BBC television series. BBC Books, £18.99.

liberation was short-lived. After the Italian government embraced the Nazi racial laws in 1938, Segrè recognized that Mussolini's Italy was no place for him.

Segrè was visiting Berkeley when the laws were promulgated. By this time, he was married and a father. Luck and circumstance combined to enable his wife, who was a German refugee, and his infant son to join him in the United States in 1939, shortly before the war started. He lived on milk shakes while he waited for their arrival, unable to stomach anything else. In 1943, at Los Alamos, J. Robert Oppenheimer told him that his mother had been captured in a Nazi roundup. Segrè recalls: "From 1930 to 1946, for over sixteen long years, my youth was dominated by the specter either of war or of political disaster. . . . I have never been able to live simply from day to day, and the anticipation of events, which unfortu-

nately often came to pass, and from which I did not know how, or could not, defend myself, has caused me much suffering."

Berkeley was a haven, but a problematic one at first because Lawrence, while generous and friendly enough to make a paying place for him, could also be volatile and exploitative. Segrè realized that his one salvation was "to do good physics", which he did, although he took pains to keep himself away from Lawrence's total authority by obtaining a teaching appointment in the physics department. The work on plutonium at Berkeley brought Segrè into collaboration with Glenn Seaborg, whom he calls a man of "unbridled personal ambitions" and an "empire builder". He compares Seaborg unfavourably with "other emperors I have known," including Fermi, who "was emperor because subjects and equals ranked him as such". Segrè declares that after the war he switched to nucleon-nucleon interactions because of Seaborg's control of the radiochemical arena.

Segrè relates few details about his actual work in physics, but does provide arresting observations on his approach to experimentation. As a young physicist in Europe, he worked with Pieter Zeeman and Otto Stern, learning thoroughness from the former and rigour from the latter, but from each "more in the philosophy of experimentation than in technical details". He also absorbed a great deal from the simplicity and directness of Fermi's physics, and not only at the laboratory bench. He takes obvious pleasure in recounting Fermi's remark on a lecture by Oppenheimer (whom Segrè found irritatingly pretentious): "Emilio, I must be

getting senile. I went to a learned theoretical seminar and could not understand anything except the last words, which were, 'And this is Fermi's theory of beta decay.'"

For many years Segrè pursued what he called "physics without apparatus", imaginatively using simple, almost rudimentary techniques to detect forbidden spectral lines, find new chemical elements and separate nuclear isomers. Once he had switched into nucleon-nucleon interactions, he entered the mode of what he calls "battleship experiments" — experiments such as the search for the antiproton that were obvious to do and could be done effectively if one had a high-energy machine and clean techniques. Of course, the techniques and apparatus were not so simple to devise, and Segrè reports that the experimental protocol required agreement among many people about how the results were to be published and credit allocated. The protocols did not satisfy the physicist Oreste Piccioni, who has claimed that Segrè and others stole his ideas, but Ernest Lawrence rejected Piccioni's complaint at the time and Segrè adds nothing to the issue in his autobiography.

A Mind Always in Motion is rich in further anecdotes and tales about, among other topics, family, physicists, politics, and even young love. It is a valuable memoir and a good read — frank, prickly, crisp and captivating, much like the man. □

Daniel J. Kevles is in the Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA.



"How cheerfully he seems to grin..." — the rare false gavia is an endangered inhabitant of the Berbak Nature Reserve on the east coast of Sumatra, Indonesia. Although classified with the Crocodylidae family, its long, narrow snout suggests that it is more closely related to the gavia (*Gavialis gangeticus*) found in parts of southern Asia. From *Wetlands in Danger: a World Conservation Atlas* edited by Patrick Dugan. Oxford University Press, \$35.

Lake zoobenthos

Annie Duncan

A Treatise on Limnology: By G. Evelyn Hutchinson, edited by Yvette Edmondson. Wiley: 1993. Pp. 944. £103, \$125.

"To everyone I say farewell"; sadly this is the last book of G. Evelyn Hutchinson and the fourth and last volume of his monumental *Treatise on Limnology* series. We have waited for this since 1975 when Volume III, *Limnological Botany*, was published. Meanwhile we have been entertained by his memoirs *The Kindly Fruits of the Earth* (Yale University Press, 1979) and instructed by his text *An Introduction to Population Ecology* (Yale University Press, 1978).

The fourth volume of the *Treatise on Limnology* adds the description of deep lake zoobenthos to the grand scheme of the series, which deals with physicochemistry, planktonic animals and plants and littoral benthos. Hutchinson had plans for a fifth volume but "it soon became clear that a properly comprehensive version of the *Treatise* would take far more time than is normally allotted to a human life". So, as his editor Yvette Edmondson comments, "it is sad to see what we have missed". The fourth volume was not completed before Hutchinson died but he was wise in his choice of an editor in 1988, and its completion has been a work of love, with Yvette Edmondson being supported in her task by his many students, colleagues and friends.

There are six chapters. The first defines types of zoobenthos which are then treated separately: the attached and unattached haptobenthos (protists, sponges, rotifers and so on), with some consideration on the differing nature of attachment surfaces; then the freely moving gastropod molluscs, followed by the aquatic insects, dealing first with juvenile stages of mayflies, Odonata and stoneflies and then with the bugs and beetles with wholly aquatic adults.

As in the previous volumes, the text is comprehensive, culling and ordering facts from a bibliography of 1,735 titles. Although these titles span the period from 1618 to 1991, over half were published between 1950 and 1980, when Hutchinson must have been reading more than one scientific paper a day. At the end of every chapter, there is a very useful summary of the main conclusions. The whole series is a grand monument to the scientific life of one man with a "sparkling mind" whose influence will continue after life even more than during it. □

Annie Duncan is in the Department of Biology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.

Gender, status and family fortunes in the white-fronted bee-eater

Stephen T. Emlen and Peter H. Wrege

Models of how people make decisions are central to economic theory. But when cast in a darwinian framework, such models provide insights into animal social behaviour. This is especially true of family-based societies in which the profitability of pursuing different behaviours is strongly influenced by gender, kinship and social position within the group. The potential of this approach is shown by modelling the reproductive decisions facing individuals in a complex bird society, that of the white-fronted bee-eater.

OPTIMALITY modelling has become a major focus in behavioural ecology¹⁻⁴, but quantitative use of decision-making theory in the study of social interactions has lagged far behind. The variability of social interactions prompted early workers to view them as outside the realm of evolutionary analysis^{5,6}. But it is now recognized that natural selection can operate on the decision-making capability underlying the choice of a behaviour, provided that heritable variation exists in the decision rules^{2,6-9}.

Decision models as developed by economists view humans as rational decision-makers who compare options and choose the one that maximizes some utility of interest¹⁰⁻¹⁴. Evolutionary biologists also view organisms as optimizing agents whose assessment abilities have been tuned by natural selection. The assumption is that the utility being maximized is fitness (survival and reproduction)^{7,15,16}, measured in terms of the number of offspring-equivalents¹⁷ produced during an individual's lifetime (its inclusive fitness¹⁸). We suggest that when inclusive fitness is influenced by social, demographic and genealogical factors, individuals will use such factors in making their proximate decisions of which reproductive tactics to follow. They will do so by assessing their relative position, and that of potential mates, in their family or other social group.

We develop a decision model that tests this hypothesis in two ways. First we calculate which of the various reproductive options are most profitable for individuals of specific social, demographic and genealogical backgrounds. This allows us to predict the choices that individuals of differing backgrounds should make; we can then compare the predictions with real-life choices.

A limitation of such analyses is the assumption that each individual experiences 'average' conditions. Therefore we also calculate individual-specific profitabilities of the different reproductive options, using the same model equations but substituting information from each individual's particular family situation. We test the observed distribution of decisions against those predicted by our model, as well as by alternative models.

Social dynamics

White-fronted bee-eaters (*Merops bullockoides*) are birds common to the savannas of East and Central Africa. They live in colonies averaging 200 individuals, and they are 'cooperative breeders', in that non-breeding individuals often help at the nests of others. The core of the society is the extended family group, typically consisting of five to nine individuals (range 2-17), and representing up to four overlapping generations¹⁹. The birds are monogamous, divorce is uncommon¹⁹, and incest is absent. Females, unlike males, leave their natal groups to join the families of their male partners in 90% of pairings²⁰. Once paired, females typically do not return to their biological families.

This social organization provides individuals with a complex array of reproductive options. Both sexes can benefit in various ways (Fig. 1). Each season, individuals must decide whether or

not to pair and breed or to help the breeding efforts of others. They may also engage in a number of secondary tactics (Fig. 1) serving either to augment fitness benefits accrued from the primary role or to offset fitness losses in the event that the primary role is unsuccessful.

The model

In applying decision theory to bee-eaters, we calculate the expected profitabilities (inclusive fitness payoffs) associated with the primary tactics of breeding and helping. The data come from five years (1980-1984) of a longitudinal study conducted in Lake Nakuru National Park, Kenya. The model comprises equations that incorporate all relevant factors currently known to affect the reproductive success (RS) of bee-eaters. They include variables for current as well as future benefits. The benefits of breeding accrue directly, through the production of offspring. The benefits of helping are accrued only indirectly in this species, through effects on the survival of non-descendent kin²¹. Variables are defined in Box 1.

The expected profitability of breeding is the sum of expected current and future direct effects (EW_c and EW_f).

$$EW_c = \left[W_b + \sum_{i=1}^n P_{h(i)} W_h \right] P_S(2r) \quad (1)$$

BOX 1 Definitions of symbols used in equations

Measures of fitness	
W_c	Current direct fitness
W_f	Future direct fitness
W_D	Total direct fitness ($W_c + W_f$)
W_{ci}	Current indirect fitness
W_{fi}	Future indirect fitness
W_i	Total indirect fitness ($W_{ci} + W_{fi}$)
W_b	RS of pair in nests that succeed
W_h	Increment in RS per successful nest due to each helper
Event probabilities	
$P_{n(i)}$	Probability that individual (i) becomes a helper (dependent on kinship and age, see Fig. 2a)
P_r	Probability of being recruited, defined as failure of a pair to initiate nesting (dependent on kinship and age of the male partner, see Fig. 2b)
P_n	Probability of breeding failure after nest initiation (dependent on kinship of the male partner and group size, see Fig. 2c)
P_S	Total probability of nesting success ($1 - (P_r + P_n - (P_r P_n))$)
P_{ra}	Probability another nest is active in the family when the primary nesting fails
P_{in}	Probability of receiving help from current offspring (W_c) in year $t+1 = I_{x(1,y)}[(I_{x(pn)}) P_{n(0.5,1)} + (I_x - I_{x(pn)}) P_{n(0.25,1)} - ((I_{x(pn)}) P_{n(0.5,1)})(I_x - I_{x(pn)}) P_{n(0.25,1)}]$
Other variables	
r	Coefficient of relatedness between ego and the young reared
t	Age
k	Group size ($k_n + 2$)
k_n	Number of helpers attending nest ($= \sum P_{n(i)}$ over all potential helpers)
I_x	Annual adult survival
$I_{x(1,y)}$	Survival to age 1 year
$I_{x(pn)}$	Survival of both members of a pair

which simplifies to: $EW_t = [\Delta W_{c(t+1)} | k_h \rightarrow k_h + w_c^*] P_{th}(2r)$, where $\Delta W_{c(t+1)}$ is the change in expected RS the following year, given that the number of helpers (k_h) is increased by the number of young fledged in the current year (W_c^* , equivalent to equation (1) before being multiplied by $2r$), devalued by P_{th} , the probability of receiving help from current offspring in year $t+1$ (see Box 1). Although equation (2) as simplified could be iterated for any number of years, additional benefits become so small in our population (<1% of the total value) that we include only a single future year in the model.

The expected profitability of helping is the sum of expected current and future indirect fitness gains (EW_{ci} and EW_{fi}). In each case, indirect fitness is measured as the increase in RS for the helped pair due to the additional aid of the helper, adjusted for the helper's relatedness to the young produced.

which simplifies to: $EW_{ci} = [\Delta W'_c | k'_h \rightarrow k'_{h+1}] 2r$.

which simplifies to: $EW_{fi} = [\Delta W'_{ci}(r+1) | k'_h \rightarrow k'_{h+w'_{ci}}] P_{th}(2r'')$, where symbols with primes indicate values for the helped pair, r is the coefficient of relatedness between the helper and the

young reared, W_{cl}^* is equivalent to equation (3) before multiplying by $2r$, and r'' is the helper's relatedness to the young reared the following year.

Because we could detect no differences in adult survival costs between breeders, helpers and non-participants, our model focuses on differential benefits. Two factors predictably influence the expected fitness gains of breeding and helping for bee-eaters.

Helper number. The average productivity of unaided breeders is essentially doubled with each additional adult provisioning the nest²². This benefit is frequency-independent: helper contribution is constant regardless of group size²². In our model, the fitness effects of helpers are realized through k_h and P_h .

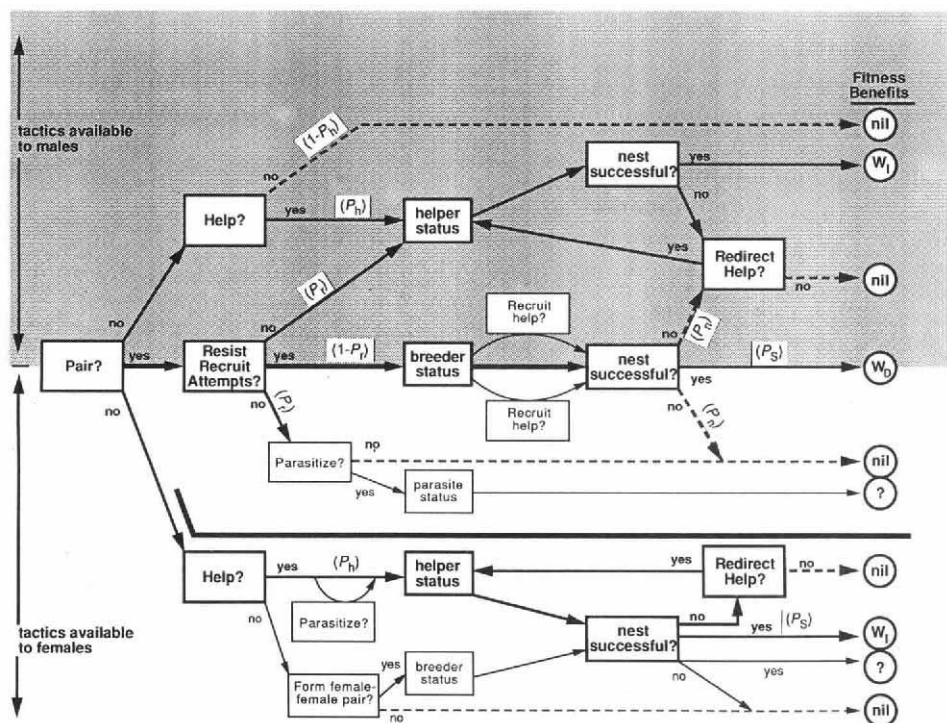
Family structure. Social, demographic and genealogical relationships between family members strongly affect the expected fitness of alternative tactics. These effects enter our model through variables P_h , P_I and P_n .

P_h , the probability that an individual becomes a helper (Fig. 2a), is significantly influenced by pair-bond status, mean kinship with the breeding pair, and age²³. Breeders with younger, unpaired, close genetic relatives are most likely to obtain the latter as helpers, thus achieving high payoffs through breeding.

P_r , the probability of being recruited (Fig. 2b; here operationally defined as any failure of paired birds to initiate nesting), commonly reflects coercive behaviour on the part of some breeders to obtain others as 'involuntary' helpers²⁴. Recruitment devalues the expected profitability of breeding for potential recruits, while augmenting it for the recruiting pair (by increasing k_h). Young, paired males with closely related, older breeders present in their families are at greatest risk of having their own breedings disrupted by recruitment²⁴.

P_n , the probability of breeding failure following nest initiation (Fig. 2c), is significantly influenced by both kinship and group size. Nest failure is high when close genetic relatives are breeding; this presumably reflects both late coercion and voluntary abandonment to redirect help to the relative's nest. Nest failure is low when group size is large, reflecting the importance of helpers in enhancing nestling survival.

FIG. 1 Flow diagram of bee-eater reproductive decisions. The primary decisions facing each individual are whether or not to pair, breed or help. Having adopted a primary role, individuals may engage in a number of secondary tactics: breeders may harass others to recruit them into becoming helpers at their nest; Failed breeders or helpers may become helpers at another nest ('redirected' helping); a female may become a conspecific brood parasite, and lay an egg in another nest at the colony; two females may form a female-female pairbond, seek extra-pair fertilizations, and together rear young without male assistance; and males may seek extra-pair fertilizations with additional females (not shown). Bottom third of the figure shows options available to females remaining in their natal families. Line thicknesses represent the relative frequency of expression; broken lines indicate that no fitness is accrued. Algebraic expressions along pathways refer to variables in the decision model: W_D , Current and future direct fitness benefits, W_I , current and future indirect fitness gains, and $?$, unmeasured but slight direct fitness accrued through intraspecific parasitism and female-female pairings.



		Age			
		a	1	2	≥3
Kinship	0.5	0.69	0.38	0.17	
	0.25	0.56	0.33	0.07	
	<0.25	0.21	0.08	0.02	

		Age			
		b	1	2	≥3
Kinship	0.5	0.48	0.23	0.09	
	0.25	0.22	0.08	0.03	
	<0.25	0.13	0.03	0.02	

		Group size			
		c	2	3	4
Kinship	0.5	0.70	0.50	0.31	
	0.25	0.57	0.37	0.20	
	<0.25	0.44	0.26	0.13	

FIG. 2 The importance of family structure for three variables used to calculate the profitability of breeding. Each matrix is based on a logistic regression model that significantly predicted the behaviour. *a*, P_h , the probability of becoming a helper, partitioned according to an individual's age and mean kinship to the breeding pair. Values pool the probabilities for paired and single birds. (Values are calculated for all individuals in the population, and thus differ from values reported elsewhere²³ where analysis was restricted to non-breeding individuals.) *b*, P_r , the probability of being recruited, partitioned by the age of the male mate and his mean kinship to the most closely related, equal-aged or older, pair in his family. Although both members of a pair may be harassed during recruitment attempts, it is the male that 'yields' to become a helper at the nest of the recruiter(s). (Values omit recruitment occurring after nest initiation, and thus differ slightly from previously published values²⁴.) *c*, P_n , the probability of breeding failure after nest initiation, partitioned according to group size tending the nest and the male's mean kinship to the most closely related other breeding pair in his family.

Decisions, decisions: males

Because male bee-eaters stay at home, they can recoup much of the loss of failed breeding attempts by becoming helpers at the nests of close relatives (Fig. 1)²². A male's expected profitability from breeding therefore includes not only his current and future direct fitness (W_D), but also a portion of the indirect fitness available from helping:

$$W_D + P_r W_{ci} + [1 - (P_s + P_r)](P_{ra})(W_{ci}) + W_{fi} \quad (5)$$

where subsequent terms represent current indirect fitness obtained if the male is recruited, if he redirects help following natural nest failure, and his future indirect fitness, respectively. Figure 3a provides the breeder payoff matrix for males based on equation (5). A comparison of cells of like kinship between matrices *a* and *b* in Fig. 3 suggests that all paired males should breed. This is largely true: 88% of such males ($n=209$ male-seasons) attempted breeding.

Because helping yields a lower payoff than breeding, it represents a second-best option for paired males. Male helpers, however, commonly occur¹⁹. Who are they? As predicted, such helpers consist almost entirely (86%) of individuals with no possibility of breeding; unpaired birds or breeders whose nesting attempts have already failed.

The services of helpers constitute a valuable resource for breeders. A breeder may attempt, by harassment, to recruit helpers beyond those that assist voluntarily. Figure 4a contrasts the net benefits (to the breeder) and costs (to the potential recruitee) of the latter yielding (helping) as opposed to resisting (continuing breeding). The net cost of yielding increases dramatically as kinship to the recruiter decreases. Figure 4b confirms the prediction that potential recruitees should resist more, the less related they are to the recruiting pair. Faced with differential resistance, breeders should preferentially harass those individuals most likely to give in: younger, closely related kin; this prediction is also supported (Fig. 4c). These results demonstrate that kinship, as well as dominance, determines the amount of leverage that one family member has over another^{24,25}.

Decisions, decisions: females

For an unpaired female, the decision to breed is essentially one of leaving home to become an in-law in the family of her mate (Fig. 1). Profitability of breeding for such a female depends on the expected RS of her mate; this depends on his 'quality', defined by his social position within his family (Fig. 3c). As unpaired females retain the option of staying at home and helping, thereby accruing fitness credits (Fig. 3b), they should be sensitive to the difference between these profitabilities.

We tested this prediction by calculating expected payoffs of the two options for each female that paired, using the model equations but substituting individual-specific information about her chosen mate and her helping options at home. Comparing the two profitabilities, we assigned each female's pairing decision as 'correct' or 'incorrect'. 'Incorrect' choices represent decisions not explained by our model. 91% of the observed pairing decisions were consistent with the family-structure hypothesis. By our calculations, only 7 of the 74 females that paired would have realized higher fitness by staying at home and helping.

We compared this observed frequency of 'errors' with those generated by simulations based on two alternative hypotheses of random choice (P.H.W. and S.T.E., manuscript in preparation). The first proposes that females pair randomly, both with respect to helping payoffs in their natal families and to the characteristics of their prospective mates. The second hypothesizes that females may assess helping payoffs at home before making a pairing decision, but then choose mates randomly without regard to male 'quality'.

Neither performed as well as the family-structure model in predicting the observed pattern of pairings. The first model generated an average of 16 'incorrect' pairings and never achieved as few as seven. The second averaged 13 errors and achieved seven or fewer less than 1% of the time (both $P < 0.01$).

These results show that unpaired females weigh-up the expected fitness consequences of helping and breeding. They choose to remain as helpers precisely when the expected payoff exceeds that of pairing and breeding.

When a female does pair, she joins the family of her mate, a group with whom she lacks kinship ties. Until she has offspring of her own, the expected profitability of helping for such a female is nil (Fig. 3b; $r=0$). As predicted, paired females rarely become helpers (only 13 of 98 potential instances²³).

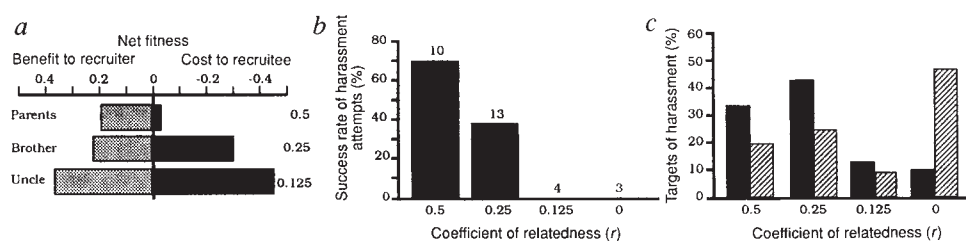
There is thus a pronounced difference between the secondary options available to paired females and males. The former are unable to become helpers to recoup lost fitness, should they fail in breeding. The model predicts conflict between members of the pair about nest abandonment, because females have much more to lose. Our data, though meagre, concur: in 8 of 9 cases where it was known that one partner remained after the other moved out, the persistent parent was female.

What additional options are there for females who fail to breed or who remain unpaired? Two tactics have been observed: the parasitic laying of an egg in the nest of a conspecific²⁶, and

a				b				c						
		Age					Age					Age		
		1	2	≥3			1	2	≥3			1	2	≥3
Kinship	0.5	0.55	0.58	0.68	Kinship	0.5	0.54			Kinship	0.5	0.24	0.38	0.55
	0.25	0.60	0.69	0.82		0.25	0.27				0.25	0.51	0.63	0.78
	0.125	0.68	0.77	0.91		0.125	0.13				0.125	0.64	0.74	0.89
	0	0.77	0.82	0.83		0	0				0	0.77	0.82	0.83

FIG. 3 Payoff matrices of the average profitabilities of breeding and helping. Cells represent different social, demographic and genealogical categories of individuals. *a*, The profitability of breeding for males, calculated using equation (5). Paired males, because they maintain kinship ties to their extended family groups, can realize indirect fitness gains should their own initial breeding attempts fail (through redirected or recruited helping). Values thus include not only direct but also applicable indirect fitness benefits. *b*, The profitability of helping (for individuals of either sex and any age), calculated using equations (3) and (4). No direct benefits accrue from helping in this species. Indirect fitness benefits are proportional to the genetic relatedness between the helper and recipient breeding pair. *c*, The profitability of breeding for paired females, calculated using equations (1) and (2). In the process of pairing, females leave their natal groups and live as in-laws in the families of their mates. Because the payoff of helping non-kin is nil, this matrix includes only direct benefits. Kinship is defined relative to the most closely related breeding pair within the family; profitabilities given in offspring equivalents¹⁷.

FIG. 4 Predicting patterns of recruitment. *a*, The net benefits and costs of successful recruitment, measured in offspring equivalents, to the recruiter and recruitee, respectively. Values are shown for three dyads of differing relatedness (labelled by identity of the recruiter). Values calculated from equations (1) to (4). *b*, The probability that harassment attempts are successful, partitioned according to kinship (defined relative to the recruiting pair). Numbers above bars denote sample sizes. The probability of successful recruitment drops precipitously as kinship decreases. *c*, Expected (hatched) and observed (solid) frequencies of harassment attempts, partitioned by relatedness between the participants. Data and definitions as in *b*.



Expected values are calculated assuming that individuals harass others randomly within their families. The null hypothesis of random targeting is rejected ($\chi^2 = 16.18$, $P < 0.01$). Breeders non-randomly select close genetic kin as targets of their harassment attempts.

the formation of female–female reproductive pairs which attempt to breed without male participation. The payoffs of both are low, and then happen only when primary options are denied to them. But they demonstrate the tremendous behavioural plasticity available to this species, and underscore the relative disadvantage to females of patrilocal residence (Fig. 1).

Discussion

Because bee-eaters live in a family-based society, kinship is a critical variable in our model. Individuals with close genetic relatives are more likely both to become helpers and, if breeding, to acquire helpers. A second key variable is social position: older individuals are more likely to get help, voluntary or otherwise, whereas younger individuals are more likely to find their breeding attempts disrupted.

Two important findings emerge from this study. First, reproductive decisions are profoundly affected by sex-biased dispersal. Females that have left their natal families largely forfeit future benefits as helpers, an option still available to family-dwelling males. This presumably explains why paired females are less likely than their mates to abandon an ongoing breeding effort, and single females are very choosy about whether, and with whom, to mate. This result should apply to other societies with extended families and long-term pair bonds, with the proviso that the pattern will be reversed when dispersal is male-biased. Second, there are circumstances in which helping is actually pre-

ferable to breeding²⁷. This runs counter to most ideas, which assume that breeding opportunities are limited and that helpers are making the best of a bad situation until a breeding vacancy occurs^{28–31}. For female bee-eaters, the relative profitabilities of helping and breeding depend, respectively, on the structure of one's own family and on the social milieu of one's potential mate. For many single females with close kin breeding at home, it is better to help than to pair and try to breed with a male of low social quality.

We began by arguing that the social dynamics of organisms living in complex societies can be analysed using decision models that incorporate inclusive fitness as their currency. Our success in predicting the behavioural options adopted by individual bee-eaters strengthens this argument. Although our example is avian, we believe that similar predictor variables of gender, kinship and social status will be important for understanding the behavioural tactics adopted by most social species. Recent studies of organisms as disparate as social insects^{32–36}, birds^{37–44} and mammals^{16,45–47} reinforce this view. We encourage behavioural ecologists and social scientists alike to incorporate these methods into their analyses of social and family dynamics of all species that live in highly structured societies. □

Stephen Emlen and Peter Wrege are at the Section of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA.

- Pyke, G. H. *A. Rev. Ecol. Syst.* **15**, 523–575 (1984).
- Stephens, D. W. & Krebs, J. R. *Foraging Theory: Princeton Monograph in Behavior and Ecology* (Princeton Univ. Press, Princeton, 1986).
- Schoener, T. W. in *Foraging Behavior* (eds Kamil, A. C., Krebs, J. R. & Pulliam, H. R.) 5–66 (Plenum, New York, 1987).
- Krebs, J. R. & Houston, A. I. in *Ecological Concepts: The Contribution of Ecology to an Understanding of the Natural World* (ed. Cherratt, J. M.) 309–338 (Blackwell Scientific, Oxford, 1989).
- Tinbergen, N. *Social Behaviour in Animals* (Methuen, London, 1953).
- Crook, J. H. in *Comparative Socioecology: the Behavioural Ecology of Humans and other Mammals* (eds Standen, V. & Foley, R. A.) 1–36 (Blackwell, Oxford, 1989).
- Dunbar, R. in *Comparative Socioecology: the Behavioural Ecology of Humans and other Mammals* (eds Standen, V. & Foley, R. A.) 131–149 (Blackwell, Oxford, 1989).
- McFarland, D. J. *Nature* **269**, 15–21 (1977).
- Dawkins, R. *The Extended Phenotype* (Freeman, Oxford, 1982).
- Becker, G. S. in *Economics of the Family: Marriage, Children, and Human Capital* (ed. Schultz, T. W.) 299–344 (Univ. of Chicago Press, Chicago, 1975).
- Becker, G. S. *The Economic Approach to Human Behavior* (Univ. of Chicago Press, Chicago, 1976).
- Edwards, W. & Tversky, A. *Decision Making* (Penguin, London, 1967).
- Mansfield, E. *Microeconomics: Theory and Applications* (Norton, New York, 1979).
- French, S. *Decision Theory* (Halstead, New York, 1986).
- McFarland, D. & Houston, A. *Quantitative Ethology: A Solid State Approach* (Pittman, London, 1981).
- Dunbar, R. I. M. *Reproductive Decisions: An Economic Analysis of Gelada Baboon Social Strategies* (Princeton Univ. Press, Princeton, 1984).
- West-Eberhard, M. J. *Q. Rev. Biol.* **50**, 1–33 (1975).
- Hamilton, W. D. *J. Theor. Biol.* **7**, 1–16 (1964).
- Emlen, S. T. in *Cooperative Breeding in Birds: Long-term Studies of Ecology and Behavior* (eds Stacey, P. & Koenig, W.) 489–526 (Cambridge Univ. Press, Cambridge, 1990).
- Hegner, R. E. & Emlen, S. T. *Ethology* **76**, 189–222 (1987).
- Emlen, S. T. & Wrege, P. H. *Behav. Ecol. Sociobiol.* **25**, 303–320 (1989).
- Emlen, S. T. & Wrege, P. H. *J. Anim. Ecol.* **60**, 309–326 (1991).
- Emlen, S. T. & Wrege, P. H. *Behav. Ecol. Sociobiol.* **23**, 305–315 (1988).
- Emlen, S. T. & Wrege, P. H. *Nature* **346**, 331–333 (1992).
- Emlen, S. T. & Wrege, P. H. *Nature* **359**, 24 (1992).
- Emlen, S. T. & Wrege, P. H. *Ethology* **71**, 2–29 (1986).
- Stacey, P. B. & Ligon, J. D. *Am. Nat.* **137**, 831–846 (1991).
- Emlen, S. T. *Am. Nat.* **119**, 29–39 (1982).
- Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton Univ., Princeton, 1987).
- Emlen, S. T. in *Behavioral Ecology: An Evolutionary Approach* 3rd edn (eds Krebs, J. & Davies, N.) 301–337 (Blackwell Scientific, Oxford, 1991).
- Koenig, W. D., Pitelka, F. A., Carmen, W. J., Mumme, R. L. & Standback, M. T. *Q. Rev. Biol.* **67**, 111–150 (1992).
- Visscher, P. W. *Behav. Ecol. Sociobiol.* **18**, 453–460 (1986).
- Page, R. E., Robinson, G. E. & Fondrik, M. K. *Nature* **338**, 576–579 (1989).
- Ratnick, F. L. W. & Visscher, P. W. *Nature* **342**, 796–797 (1989).
- Mueller, U. G. *Science* **254**, 442–444 (1991).
- Reeve, H. K. & Nonacs, P. *Nature* **359**, 823–825 (1992).
- Komdeur, J. *Nature* **358**, 493–495 (1992).
- Reyer, H. U. *Behaviour* **96**, 278–303 (1986).
- Stacey, P. B. & Ligon, J. D. *Am. Nat.* **130**, 654–676 (1987).
- Mumme, R. L., Koenig, W. D. & Pitelka, F. A. *Am. Nat.* **131**, 654–677 (1988).
- Curry, R. L. *Behav. Ecol. Sociobiol.* **22**, 141–152 (1988).
- Zack, S. W. & Rabenold, K. N. *Anim. Behav.* **38**, 235–247 (1989).
- Rabenold, P. P., Rabenold, K. N., Piper, W. H., Haydock, J. & Zack, S. W. *Nature* **348**, 538–540 (1990).
- Davies, N. B. *Dunnock Behaviour and Social Evolution* (Oxford Univ. Press, Oxford, 1992).
- Noe, R. & Sluiter, A. A. *Behaviour* **113**, 117–170 (1990).
- Creel, S. R. & Waser, P. W. *Behav. Ecol.* **2**, 7–15 (1991).
- Packer, C., Gilbert, D. A., Pusey, A. E. & O'Brien, S. J. *Nature* **351**, 562–565 (1991).

ACKNOWLEDGEMENTS. We thank C. E. McCulloch and D. W. Winkler for analytical advice and N. B. Davies, R. Dunbar, J. R. Krebs, C. Packer, D. W. Winkler and especially N. J. Demong for comments on the manuscript. The field research was supported by the NSF, the NGS and the John Simon Guggenheim Foundation with permission from the government of Kenya.

Potential impact of climate change on world food supply

Cynthia Rosenzweig* & Martin L. Parry†

* Columbia University and Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA

† Environmental Change Unit, Oxford University, 1a Mansfield Road, Oxford OX1 3TB, UK

A global assessment of the potential impact of climate change on world food supply suggests that doubling of the atmospheric carbon dioxide concentration will lead to only a small decrease in global crop production. But developing countries are likely to bear the brunt of the problem, and simulations of the effect of adaptive measures by farmers imply that these will do little to reduce the disparity between developed and developing countries.

RECENT research assessing the potential effects of climate change on agriculture has focused on regional and national evaluations¹⁻³. These efforts have, for the most part, treated each region or nation in isolation, without relation to changes in agricultural production elsewhere. Recent work^{4,5} emphasizes the important role of international trade in the adjustment of the world food system to climate change-induced changes in crop yields. Crop growth models have been used in an extensive international collaboration⁶ to determine the effect of various climate change scenarios on crop yields for individual countries and geographical regions (Fig. 1). In this article we combine the data from these individual studies to obtain a global picture of the simulated change in crop yield associated with the different climate change scenarios. We then use a world food trade model to simulate the economic consequences of these potential changes in crop yields; we estimate changes in world food prices, and in the number of people at risk of hunger (defined as a measure of food energy availability—which depends on income and food price levels—relative to nutritional requirements) in developing countries.

The major finding of the study is that there appears to be a large disparity in agricultural vulnerability to climate change between developed and developing countries. This occurs even though simulated global agricultural production of major grain crop declines are only small to moderate under the climate change conditions tested. The analysis included the combined effects of climate change and increasing CO₂ on crop yields and water use. Although projected temperature change in low latitudes (where many developing countries are located) tends to be lower than the global average in the general circulation model (GCM) scenarios tested, modelled yield changes are primarily negative there, in contrast to predominantly positive yield changes in middle and high latitudes where many developed countries are located. This result has significant implications for potential future distributional aspects of the world food system.

Studies such as this explore the sensitivity of important human systems (in this case world food supply), as currently understood, to projected levels of global climate change. Such studies are initial demonstrations of the comprehensive, interdisciplinary research needed to improve understanding of the interactive biophysical and socio-economic effects that may result from global environmental change.

Climate change scenarios

Scenarios were developed from climate conditions predicted by three GCMs for doubled atmospheric CO₂ levels (Table 1). The temperature changes of these GCM scenarios (4.0–5.2 °C) are near the upper end of the range (1.5–4.5 °C) projected for doubled CO₂ warming by the IPCC^{7,8}. Mean monthly changes in temperature, precipitation and solar radiation from the appro-

priate GCM gridbox were applied to observed daily climate records (1951–80, or as many years of daily climate records as available) to create climate change scenarios for each study site.

Because atmospheric concentrations of other greenhouse gases besides CO₂ (for example, methane (CH₄), nitrous oxide (N₂O) and the chlorofluorocarbons (CFCs)) are also increasing, an 'effective CO₂ doubling' has been defined as the combined radiative forcing of all greenhouse gases having the same forcing as doubled CO₂, usually taken to be ~600 p.p.m. CO₂ level is important when estimating potential impact on crops because crop growth and water use have been shown to benefit from increased levels of CO₂ (refs 9, 10). For the crop model simulations, climate changes from the doubled CO₂ GCM simulations were used with an associated level of 555 p.p.m. CO₂; these conditions are assumed to occur in AD2060. The 555 p.p.m. level is based on the GISS GCM trace gas scenario A (ref. 11), in which the simulated climate had warmed to the effective doubled CO₂ level of ~4 °C by AD2060.

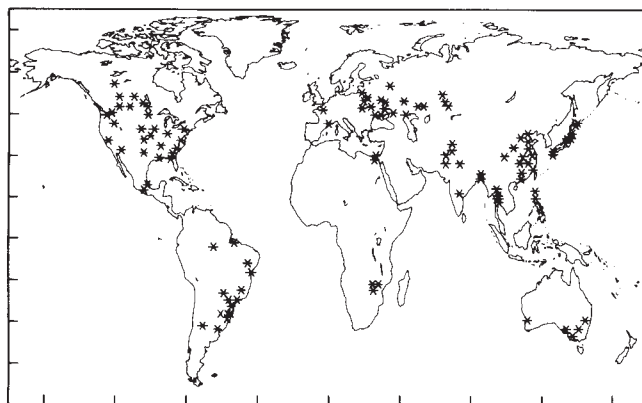


FIG. 1 Crop model sites⁶. Countries and participants were as follows: Argentina, O. E. Sala and J. M. Paruelo; Australia, B. D. Baer, W. S. Meyer and D. Erskine; Bangladesh, Z. Karim, M. Ahmed, S. G. Hussain and Kh. B. Rashid; Brazil, O. J. F. de Siqueira, J. R. B. Farias and L. M. A. Sans. Canada, M. Brklacich, R. Stewart, V. Kirkwood and R. Muma; China, Z. Jin, D. Ge, H. Chen, J. Fang and X. Zheng; Egypt, H. M. Eid; France, R. Delécolle, D. Ripoche, F. Ruget and G. Gosse; India, D. G. Rao; Japan, H. Seino; Mexico, D. Liverman, M. Dilley, K. O'Brien and L. Menchaca; Pakistan, A. Qureshi and A. Iglesias; Philippines, C. R. Escaño and L. Buendia; Thailand, M. L. C. Tongyai; Russia, G. Menzhulin, L. Koval and A. Badenko; USA, C. Rosenzweig, B. Curry, T.-Y. Chou, J. Ritchie, J. Jones and R. Peart; Uruguay, W. E. Baethgen; Zimbabwe, P. Muchena.

TABLE 1 GCM doubled-CO₂ climate change scenarios

GCM	Year*	Resolution (lat. × long.)	CO ₂ (p.p.m.)	Change in average global temperature (°C)	Change in average global precipitation (%)
GISS†	1982	7.83° × 10°	630	4.2	11
GFDL‡	1988	4.4° × 7.5°	600	4.0	8
UKMO§	1986	5.0° × 7.5°	640	5.2	15

* When calculated.

† Goddard Institute for Space Studies²⁰.‡ Geophysical Fluid Dynamics Laboratory²¹.§ United Kingdom Meteorological Office²².

Crop yield change methods

Agricultural scientists in 18 countries estimated potential changes in national grain crop yields using compatible crop models and the GCM scenarios at 112 sites (Fig. 1). The crop model linkages were developed by the US Agency for International Development¹². Simulations were carried out in regions representing 70–75% of the current world production of wheat, maize and soybean; rice production was less well represented (48% of current world production).

Site-specific estimates of yield changes were aggregated to national levels based on current regional production. The regional yield estimates represent the current mix of rainfed and irrigated production, and today's crop varieties, nitrogen management and agricultural soils. The national crop yield changes were then extrapolated to provide estimates of yield changes for the other countries and crops included in the world food trade model. National yield changes of other crops and commodity groups and regions not simulated were estimated based on three criteria: (1) similarities to modelled crops and growing conditions; (2) results from ~50 previously published and unpublished regional climate change impact studies; and (3) projected temperature and precipitation changes from the GCM scenarios. The primary sources of uncertainty in the estimates lies in the sparseness of the crop modelling sites and the lack of explicitly modelled yield changes for subsistence crops such as millet and cassava, which may respond differently to both climate change and increases in CO₂.

Estimates were made of yield changes with and without the direct physiological effects of CO₂ on crop growth, that is increased rates of net photosynthesis and reduced stomatal openings as reported from experimental results¹³. The photosynthesis ratios (555/330 p.p.m. CO₂) for soybean, wheat and rice, and maize were 1.21, 1.17, and 1.06, respectively. Changes in stomatal resistance were based on experimental results¹⁴ (49.7/34.4 s m⁻¹ for C3 crops, 87.4/55.8 s m⁻¹ for C4 crops). This method of simulating the physiological CO₂ effects on crops may provide a positive bias to yield estimates, as plants grown in experimental settings are often subject to fewer environmental and competitive stresses than are likely to be encountered in farmers' fields.

GCM scenarios and direct CO₂ effects

Figure 2 shows estimated potential changes in average national grain crop yields for the GISS GFDL and UKMO doubled-CO₂ climate change scenarios (Table 1) with and without physiological (direct) CO₂ effects on plant growth. The maps are created from nationally averaged yield changes for wheat, rice, coarse grains and protein feed estimated for each country or group of countries. When climate change is considered without direct CO₂ effects on crop growth and water use, averaged national crop yields declined everywhere, although reductions were less at middle and high latitudes. In the simulations with direct CO₂ effects, yields were positive at middle and high latitudes, and negative at low latitudes for the GISS and GFDL scenarios

which produced yield changes ranging from +30 to -30%. The UKMO scenario caused average national crop yields to decline almost everywhere (up to -50% in Pakistan).

Several factors contributed to the latitudinal differences in simulated yields. At some sites near the high-latitude boundaries of current agricultural production, increased temperatures benefited crops otherwise limited by cold temperatures and short growing seasons. In many middle and high latitude areas, where current temperature regimes tend to be cooler, increased temperatures exerted a negative influence on yields through shortening of crop development stages, but did not significantly increase heat or water stress levels. In these regions beneficial CO₂ effects dominated. The climate-change-induced warming at low latitudes brought not only accelerated growing periods for crops, but also greater heat and water stress, resulting in greater yield decreases than at higher latitudes, despite beneficial CO₂ direct effects.

Farm-level adaptations

In each participating country, the agricultural scientists used the crop models to test possible responses to the worst climate-change scenario (this was usually, but not always, the UKMO scenario; adaptation simulations were done with all three GCMs at some sites). These adaptations included changes in planting date, variety and crop, and applications of irrigation and fertilizer. Irrigation simulations assumed automatic irrigation to field capacity when plant available water dropped to 50% and 100% irrigation efficiency. These optimistic assumptions imply that water supply for irrigation would be fully available at all locations under climate change conditions. All adaptation possibilities

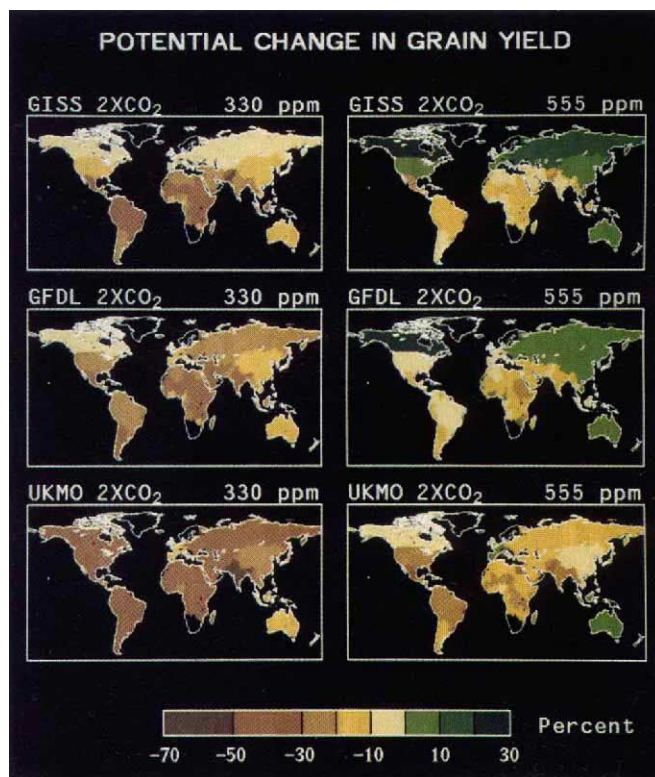


FIG. 2 Estimated change in average national grain yield (wheat, rice, coarse grains, and protein feed) for the GISS, GFDL, and UKMO climate change scenarios. The left-hand column shows the direct physiological effect on grain yield of current (330 p.p.m.) CO₂ concentration. The right-hand column shows the direct physiological effect of 555 p.p.m. Results shown are averages for countries and groups of countries in the Basic Linked System (BLS) world food trade model; regional variations within countries are not reflected.

ies were not simulated at every site and country: choices of adaptations to be tested were made by the participating scientists, based on their knowledge of current agricultural systems.

For the economic analysis, crop model adaptation results were grouped into two levels of adaptation. Level 1 implies little change to existing agricultural systems, reflecting responses to a changing climate that should be easily available to individual farmers. Level 1 adaptations included shifts in planting date (± 1 month) that do not imply major changes in crop calendar, additional application of irrigation water to crops already under irrigation, and changes in crop variety to currently available varieties more adapted to the altered climate.

Adaptation level 2 implies more substantial change to agricultural systems, possibly requiring resources beyond the farmers' means, investment in regional and national agricultural infrastructure, and policy changes, although these types of changes were beyond the scope of the analysis. Level 2 adaptations included large shifts in planting date (>1 month), increased fertilizer application (included here because of implied costs for farmers in developing countries), installation of irrigation systems, and development of new varieties (tested by manipulation of genetic coefficients in crop models). Level 2 represents a fairly optimistic assessment of world agriculture's response to the changed climate conditions tested.

To extend the adaptation site results to national yield change estimates for all the countries in the world food trade model and to the other GCM scenarios, a simplifying assumption was made based on the crop modelling results to halve the negative impact if adaptations partially compensated for the negative effects of

climate change, and to set yield changes to zero if compensation was full. If yield changes were positive, adaptation to produce even greater yield increases was not included, with the assumption that farmers would lack incentive to adapt further. This unrealistic assumption tends to underestimate potential disparities between countries able to improve productivity under the climate change scenarios and countries that cannot fully adapt. The adaptation estimates were developed only for the scenarios that included direct physiological CO_2 effects as these were judged to be most realistic.

improves realism, but many critical uncertainties remain. Although crop models allow testing of some potential improvements in agricultural production, they do not include yield-enhancing technological developments induced by negative climate change impacts. Level of adoption and efficacy of adaptive practices are also uncertain. There may be social or economic reasons why farmers are reluctant to implement adaptation measures, for example, increased fertilizer application and improved seed stocks may be capital-intensive and/or not suited to indigenous agricultural strategies. Furthermore, such measures may not necessarily result in sustainable production increases (for example, irrigation may eventually lead to soil salinization).

Yield estimates for the two adaptation levels are shown in Fig. 3. Level 1 adaptation compensated incompletely for the climate change scenarios, particularly in the developing countries. For the GISS and GFDL scenarios, level 2 adaptation compensated almost fully for negative climate change impacts. With the high level of global warming projected by the UKMO climate change scenario, neither level 1 nor level 2 adaptation fully overcame the negative climate change effects on crop yields in most countries, even when direct CO_2 effects are taken into account.

World food trade model

The basic linked system (BLS) consists of a set of linked national agricultural sector models¹⁵. It is comprised of 16 national (including the EU) models with a common structure, 4 models with country-specific structure and 14 regional group models. The political changes as well as changes in national boundaries of the very recent past are not in the BLS, although the model formulation has been adjusted, away from centrally planned economies to more market-oriented behaviour. The 20 models in the first two groups cover $\sim 80\%$ of world agricultural production; the remaining 20% is covered by 14 regional models for countries which have broadly similar attributes (for example African oil-exporting countries, Latin American high-income exporting countries, Asian low-income countries). The BLS is a general equilibrium model system, with representation of all economic sectors, empirically estimated parameters and no unaccounted supply sources or demand sinks. Countries are linked through trade, world market prices and financial flows.

The BLS does not incorporate any climate relationships *per se*. Effects of changes in climate were introduced to the model as changes in the average national or regional yield per commodity as described above. Internal economic adjustments occur as increased agricultural investment, reallocation of agricultural resources according to economic returns (including crop switching), and reclamation of additional arable land as a response to higher commodity prices. Improvements in agricultural technology are represented by annual yield trends for developed and developing countries based on historical trends. The BLS contains yield-fertilizer production functions to capture the effects that changes in fertilizer prices and subsequent changes in fertilizer applications may have on yields at the national level. The economic adjustments to climate change simulated by the BLS are assumed not to alter the basic structure of yield-fertilizer functions, even though some of these relationships (for example, yield responses to nitrogen fertilization) may be altered in a changed climatic regime and under elevated CO_2 conditions.

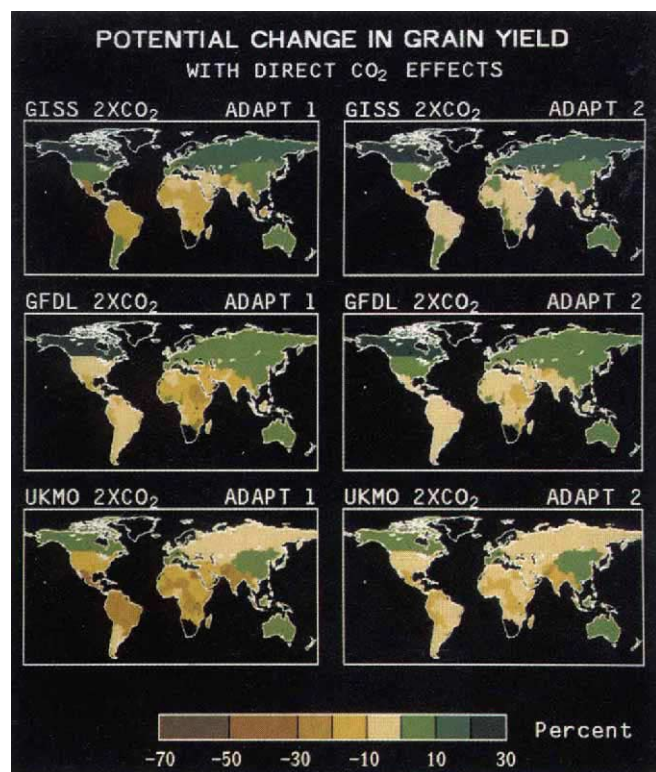


FIG. 3 Estimated changes in average national grain yield (wheat, rice, coarse grains, and protein feed with direct 555 p.p.m. CO_2 effects) under two levels of adaptation for the GISS, GFDL and UKMO doubled- CO_2 climate change scenarios⁸. Adaptation level 1 signifies minor changes to existing agricultural systems; adaptation level 2 signifies major changes. Results shown are averages for countries and groups of countries in the BLS world food trade model; regional variations within countries are not reflected.

Summary indicators of the world food system's sensitivity to the climate change scenarios include world cereal production, world cereal prices and population in developing countries (excluding China) at risk of hunger. The 'risk of hunger' indicator in the BLS was developed from estimates of the number of undernourished people in developing countries made by the FAO¹⁶; a cross-country regression was estimated, explaining the number of people at risk of hunger by a measure of food energy availability relative to nutritional requirements¹⁵. Food availability, in turn, depends on income and price levels. Average and marginal budget shares of consumption categories besides food (for example, housing and clothing) are included in the analysis. We include the risk of hunger indicator to show the possible trends in future food security, realizing that more comprehensive measures incorporating other socioeconomic variables may be devised.

Here we limit our analysis to results relating to the major cereal food crops, even though the BLS explicitly represents efficiency of feed use and trends in production and consumption of alternative livestock commodities. Thus, the reference case simulates the shift to higher-efficiency feed conversion (poultry over beef) occurring in some developed countries. Beyond the indirect effects of changes in feed costs, livestock production is a significant component of the global food system that is potentially sensitive to climatic change because of changes in range-land and animal productivity.

World food trade model scenarios

The reference scenario (a future without climate change). The reference scenario projects the agricultural system to the year 2060 with no climate change and no major changes in the political or economic context of world food trade (Table 2). Population growth rates were exogenously specified from the medium projections of the United Nations to 2025, and from World Bank projections thereafter, resulting in ~10.3 billion people by 2060 (refs 17, 18).

Economic growth rates in the BLS are endogenously determined in most of the national models, yielding a moderate projection of world economic growth for the reference scenario. A 50% trade liberalization in agriculture (for example, removal of import restrictions) is introduced gradually by 2020. The analysis of trade liberalization is restricted to removal of distortions between trade prices and domestic prices at the level of agricultural raw materials.

Technology is projected to increase yields over time, but at a slowing rate based on historical trends. The rate of exogenous

Growth rate %	1980–2000	2000–2020	2020–2040	2040–2060
Population	1.7	1.3	0.8	0.5
GDP	2.9	2.0	1.5	1.1
Cereal yield	1.2	0.7	0.5	0.4
Agricultural production	1.8	1.3	1.0	0.7

All growth rates refer to world average annual per cent growth during the indicated period.

technical progress starts from historical values (1.3% in the 1980s) and for cereal crops approaches 0.5% per annum by 2060. Availability of arable land for expansion of crop production is based on FAO¹⁹ data.

Climate change scenarios. The food trade simulations for the three GCM scenarios were started in 1990. The yield changes estimated from the crop model simulations were applied linearly up to 2060 to the yields simulated by the BLS, which include the effect of technological improvement. Although the testing of climate change impacts without farm-level adaptation is unrealistic, it is done for the purpose of establishing a baseline with which to compare the effects of farmer response.

Scenarios including the effects of farm-level adaptations. The next step in the analysis involved adjusting the climate-induced yield changes assuming adaptation levels 1 and 2 described above. We can safely assume that at least some farm-level adaptations will be adopted, especially techniques similar to those tested in adaptation level 1. Policy, cost and water resource availability were assumed not to be barriers to adaptation.

World food trade results

The reference scenario. Assuming that population growth, economic growth, technological progress and trade liberalization proceed as specified above without climate change, world cereal production (wheat, rice, maize, millet, sorghum and minor grains) is estimated to grow to 3,286 million metric tons (m.m.t.) in year 2060 (compared with 1,795 m.m.t. in 1990). Cereal production in developing countries grows to exceed production in developed countries by 2020. Despite slowing gains in yield increases, food production (measured as net calories produced) is projected to exceed population growth throughout the simulation period of the reference scenario.

Cereal prices are estimated at an index of 121 (1970 value, 100) for the year 2060, reversing the falling trend of real cereal prices over the past 100 years. The standard reference scenario has two phases of price development. During 1980 to 2020, while trade barriers and protection are still in place but are being reduced, there are increases in relative prices. This occurs in the short-to-medium term because a removal of subsidies leads to lower farm-gate prices and therefore disincentives to production, while consumers benefit from somewhat lower retail prices, an incentive to demand. In the longer term, price decreases follow due to efficiency gains and technical progress. The number of people at risk of hunger is estimated at ~640 million or ~6% of total population in 2060 (compared with 530 million in 1990, ~10% of total current population).

Climate change scenarios. Without direct CO₂ effects on crop yields, world cereal production is reduced by 11 to 20%, and their inclusion brings yield decreases to between 1 and 8% (Fig. 4). The world production changes mask a disparity in response to climate change between developed and developing countries (Fig. 5). The largest negative changes occur in developing regions, though the extent of decreased production varies greatly by country depending on the projected climate. By contrast, in developed countries, production is estimated to increase under all but the UKMO scenario.

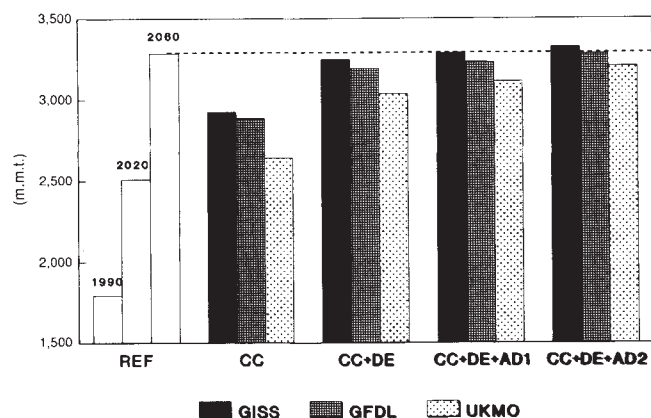


FIG. 4 World cereal production projected by the BLS for the reference, GISS, GFDL and UKMO doubled-CO₂ climate change scenarios, with (CC+DE) and without (CC) direct CO₂ effects on crop yields, and with adaptation levels 1 and 2 (AD1 and AD2). Adaptation level 1 implies minor changes to existing agricultural systems; adaptation level 2 implies major changes. (m.m.t., million metric tons).

Price increases resulting from climate-induced decreases in yield are estimated to range between ~24–145% (Fig. 5). These increases in price affect the number of people at risk of hunger. Their estimated number increases ~1% for each 2–2.5% increase in prices (depending on climate change scenario). People at risk of hunger increase by 10% to almost 60% in the scenarios tested, resulting in an estimated increase of between 60 million and 350 million people in this condition (above the reference scenario projection of 640 million) by 2060.

Scenarios including the effects of farm-level adaptations. Globally, both minor and major levels of adaptation help restore world production levels (when CO₂ effects are included), compared to the climate change scenarios with no adaptation (Fig. 4). Averaged global cereal production decreases by up to ~160 m.m.t. (0 to -5%) from the reference scenario projection of 3,286 m.m.t. with minor level 1 adaptations. With adaptations implying major changes, global cereal production responses range from a slight increase of 30 m.m.t. to a slight decrease of ~80 m.m.t. (+1% to -2.5%).

Level 1 adaptation largely offsets the negative climate change yield effects in developed countries, improving their comparative advantage in world markets (Fig. 5). In these regions, cereal production increases by 4 to 14% over the reference scenario. However, developing countries are estimated to benefit little from this level of adaptation (-9 to -12% change in cereal production). More extensive adaptation virtually eliminates global negative cereal yield impacts derived under the GISS and

GFDL climate scenarios, and reduces impacts under the UKMO scenario to one third.

Under adaptation level 1, price increases range from 10 to 100% (Fig. 5). Under adaptation level 2, cereal price responses range from a decline of ~5% to an increase of 35%. As a consequence of climate change and adaptation level 1, the number of people at risk of hunger increases by ~40 million to 300 million (6–50%) from the reference scenario of 641 million (Fig. 5). With a more significant amount of adaptation by farmers, the number of people at risk of hunger is altered by between -12 million for the GISS scenario and 120 million for the UKMO scenario (-2% and +20%). These results indicate that, except for the GISS scenario under adaptation level 2, the simulated farm-level adaptations did not mitigate entirely the negative effects of climate change on the number of people at risk of hunger, even when economic adjustment, that is, the production and price responses of the world food system, are taken into account.

Discussion

Several major points emerge from this study. Climate change scenarios near the high end of the IPCC range of doubled-CO₂ warming exerted (in most cases) a slight-to-moderate negative effect on simulated world cereal production, even when the beneficial direct effects of CO₂, farm-level adaptations and future technological yield improvements were taken into account. The only scenario that increased global cereal production was one involving major, and possible costly, changes in current agricul-

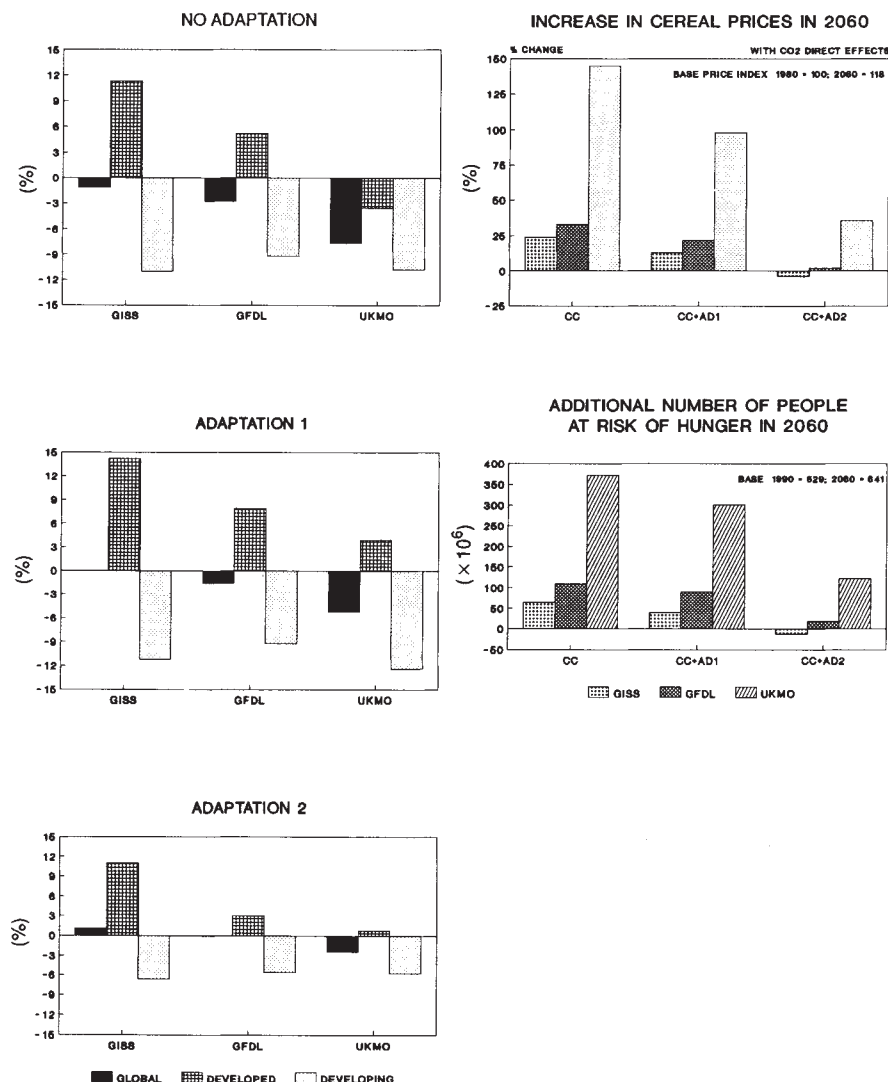


FIG. 5 Change in global, developed country, and developing country cereal production (with direct CO₂ effects), cereal prices, and people at risk of hunger in 2060 for climate change scenarios (CC), and with adaptation levels 1 and 2 (AD1 and AD2). Reference scenario for 2060 assumes no climate change, and projects global cereal production to be 3,286 m.m.t., developed country production to be 1,449 m.m.t., and developing country production to be 1,836 m.m.t.

tural systems, for example, installation of irrigation. Availability of irrigation water, however, was not included explicitly in the study; supplies may be limited under climate change conditions not only for expanding irrigation but for maintaining the current extent of irrigation in some areas. Although the overall results of the study are relatively benign, they depend strongly on the full realization in the field of beneficial direct physiological CO₂ effects on crop growth and water use as currently measured in experimental settings.

Climate change was found to increase the disparities in cereal production between developed and developing countries.

Received 9 July; accepted 2 December 1993.

1. Parry, M. L., Carter, T. R. & Konijn, N. T. (eds) *The Impact of Climatic Variations on Agriculture Vols 1 & 2* (Kluwer, Dordrecht, 1988).
2. Adams, R. M. et al. *Nature* **345**, 219–224 (1990).
3. Intergovernmental Panel on Climate Change *Climate Change: The IPCC Impacts Assessment* (eds Tegtart, W. J. McG., Sheldon, G. W. & Griffiths, D.C.) (Aust. Govt Publishing Serv., Canberra, 1990).
4. Kane, S., Reilly, J. & Tobey, J. Report AER-No. 647 (US Dept of Agric. Econ. Res. Serv., Washington DC, 1991).
5. Reilly, J. & Hohmann, N. *Am. econ. Ass. Pap. Proc.* **83**, 306–312 (1993).
6. Rosenzweig, C. & Iglesias, A. (eds) *Implications of Climate Change for International Agriculture: Crop Modeling Study* (US Environmental Protection Agency, Washington DC, in the press).
7. *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) (Cambridge Univ. Press, 1990).
8. *IPCC Climate Change 1992* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) (Cambridge Univ. Press, 1992).
9. Acock, B. & Allen, L. H. Jr in *Direct Effects of Increasing Carbon Dioxide on Vegetation* (eds Strain, B. R. & Cure, J. D.) 33–97 (US Dept. of Energy DOE/ER-0238, Washington DC, 1985).
10. Cure, J. D. & Acock, B. *Agric. For. Met.* **38**, 127–145 (1986).
11. Hansen, J., *J. geophys. Res.* **93**, 9341–9364 (1988).
12. International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) *Decision*

Whereas production in the developed world benefited from climate change, production in developing nations declined. Adaptation at the farm-level did little to reduce the disparities, with the developing world suffering the losses. Cereal prices, and thus the population at risk of hunger, increased despite adaptation. Even a high level of farm-level adaptation in the agricultural sector did not entirely prevent such negative effects. Thus, while some countries in the temperate zones may reap some benefit from climate change, many countries in the tropical and subtropical zones appear more vulnerable to the potential impacts of global warming. □

Support System for Agrotechnology Transfer Version 2.1 (DSSAT V2.1) (Dept Agron. & Soil Sci., Coll. Trop. Agric. & Hum. Resources, Univ. Hawaii, Honolulu, 1989).

13. Peart, R. M., Jones, J. W., Curry, R. B., Boote, K. & Allen, L. H. Jr in *The Potential Effects of Global Climate Change on the United States*. Appendix C Vol. 1 (eds Smith, J. B. & Tirpak, D. A.) (US Envir. Protection Ag., Washington DC, 1989).
14. Rogers, H. H., Bingham, G. E., Cure, J. D., Smith, J. M. & Surano, K. A. *J. envir. Qual.* **12**, 569–574 (1983).
15. Fischer, G., Frohberg, K., Keyzer, M. A. & Parikh, K. S. *Linked National Models: A Tool for International Food Policy Analysis* (Kluwer, Dordrecht, 1988).
16. Food and Agriculture Organization *Fourth World Food Survey*. (UN, Rome, 1984); *Fifth World Food Survey* (UN, Rome, 1987).
17. *United Nations World Population Prospects 1988* (UN, New York, 1989).
18. International Bank for Reconstruction and Development/World Bank *World Population Projections* (Johns Hopkins Univ. Press, Baltimore, 1990).
19. AGROSTAT/PC (Food and Agricultural Organization, United Nations, Rome, 1991).
20. Hansen, J. et al. *Mon. Weath. Rev.* **111**(4), 609–662 (1983).
21. Manabe, S. & Wetherald, R. T. *J. atmos. Sci.* **44**, 1211–1235 (1987).
22. Wilson, C. A. & Mitchell, J. F. B. *J. geophys. Res.* **92**(13), 315–343 (1987).

ACKNOWLEDGEMENTS. We acknowledge all our collaborators, and especially thank the IBSNAT crop modelling team for worldwide scientific and technical support, G. Fischer and K. Frohberg for their work with the BLS, and A. Iglesias for development of yield estimation techniques. We also thank C. Shashkin and R. Goldberg for project support. This work was funded by the US Environmental Protection Agency and the US Agency for International Development.

Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I

Christopher D. Lima^{*}, James C. Wang[†] & Alfonso Mondragón^{*‡}

^{*} Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, Illinois 60208, USA

[†] Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

The three-dimensional structure of the 67K amino-terminal fragment of *Escherichia coli* DNA topoisomerase I has been determined to 2.2 Å resolution. The polypeptide folds in an unusual way to give four distinct domains enclosing a hole large enough to accommodate a double-stranded DNA. The active-site tyrosyl residue, which is involved in the transient breakage of a DNA strand and the formation of a covalent enzyme–DNA intermediate, is present at the interface of two domains. The structure suggests a plausible mechanism by which *E. coli* DNA topoisomerase I and other members of the same DNA topoisomerase subfamily could catalyse the passage of one DNA strand through a transient break in another strand.

THE DNA topoisomerases are ubiquitous enzymes that can alter the topology of DNA by transiently breaking one or two strands of DNA, passing a single- or double-stranded DNA through the break, and finally resealing the break (for reviews, see refs 1–3). These enzymes are involved in a number of crucial cellular processes, including replication, transcription and recombination, and members of the DNA topoisomerase family have been identified as the targets of antimicrobial and anticancer therapeutics^{4–6}.

Escherichia coli DNA topoisomerase I, the first topoisomerase

identified⁷, is the most widely studied member of a subfamily of type-I DNA topoisomerases which includes *E. coli* DNA topoisomerase III (refs 8, 9), *Saccharomyces cerevisiae* topoisomerase III (refs 10, 11), and the archaeobacterium *Sulfolobus acidocaldarius* reverse gyrase^{12,13}. The enzyme catalyses the following reactions: relaxation of negatively supercoiled DNA, interconversion of unknotted and knotted single-stranded DNA rings, catenation and knotting of double-stranded DNA (provided that a gap or nick is present in at least one of the rings) and linking of two complementary single-stranded DNA rings into a double-stranded ring^{2,14}. All these reactions occur as a result of the transient breakage of one DNA strand to allow passage of

[‡] To whom correspondence should be addressed.

another DNA strand, or in some cases of a duplex DNA segment, through the transiently broken or gated strand, hence the classification of the enzyme as a type-I DNA topoisomerase¹⁻³. In contrast to type-II enzymes, the reactions catalysed by *E. coli* DNA topoisomerase I and all other type-I enzymes are not coupled to binding or hydrolysis of ATP. Unlike other enzymes involved in complex DNA rearrangements, such as γ/δ resolvase or phage Mu transposase¹⁵, *E. coli* DNA topoisomerase I apparently works as a monomer. The mechanism by which a molecule catalyses the concerted steps of DNA breakage, strand passage, and strand rejoining poses interesting questions.

Intact *E. coli* DNA topoisomerase I is a 97K metalloprotein containing three Zn(II) atoms in three tetracysteine motifs near the carboxy-terminus of the protein¹⁶. The DNA cleavage reaction involves transesterification between a DNA phosphodiester bond and a tyrosyl residue (Tyr 319) in the enzyme¹⁷. The reaction yields a covalent intermediate in which the tyrosine is linked to the 5' phosphoryl end of the severed DNA strand. Subsequent nucleophilic attack of the O₄-phosphotyrosine linkage by the 3'-OH of the severed DNA strand rejoins the DNA strand and frees the enzyme. The tetracysteine motifs are not required for the cleavage step; in the absence of zinc the enzyme cannot relax negatively supercoiled DNA but it can cleave single-stranded DNA¹⁸. We have previously reported the overexpression, purification and crystallization of a 67K N-terminal frag-

ment of the *E. coli* enzyme¹⁹. This fragment is incapable of relaxing supercoiled DNA but can cleave single-stranded oligonucleotides. The substrate specificity and reaction characteristics of the fragment are similar to those of the intact or the zinc-depleted protein^{18,19}. Here we report the crystal structure of this 67K fragment to 2.2 Å resolution. The structure offers a number of clues on how this subfamily of type-I DNA topoisomerases accomplish topological transformations of DNA.

Structure determination

The crystal structure was initially determined to 3.0 Å by multiple isomorphous replacement (MIR) phasing using three heavy-atom derivatives (Table 1 and Fig. 1). An initial polyaniline model was built into a solvent-flattened electron density map, using the program O aided by the Bones and Lego options²⁰. The main-chain connectivity was unclear in several regions and only 80% of the residues could be traced. Wherever possible, side chains were placed in the electron density map and the model was refined using PROLSQ (ref. 21). Refinement was continued with a 2.11 Å dataset collected at the Stanford Synchrotron Radiation Laboratory using a MAR phosphor-image plate scanner. The resolution was extended stepwise to 2.2 Å using XPLOR (ref. 22). The final model consists of 553 amino acids (4,412 atoms), or 93% of the polypeptide, and 267 water molecules.

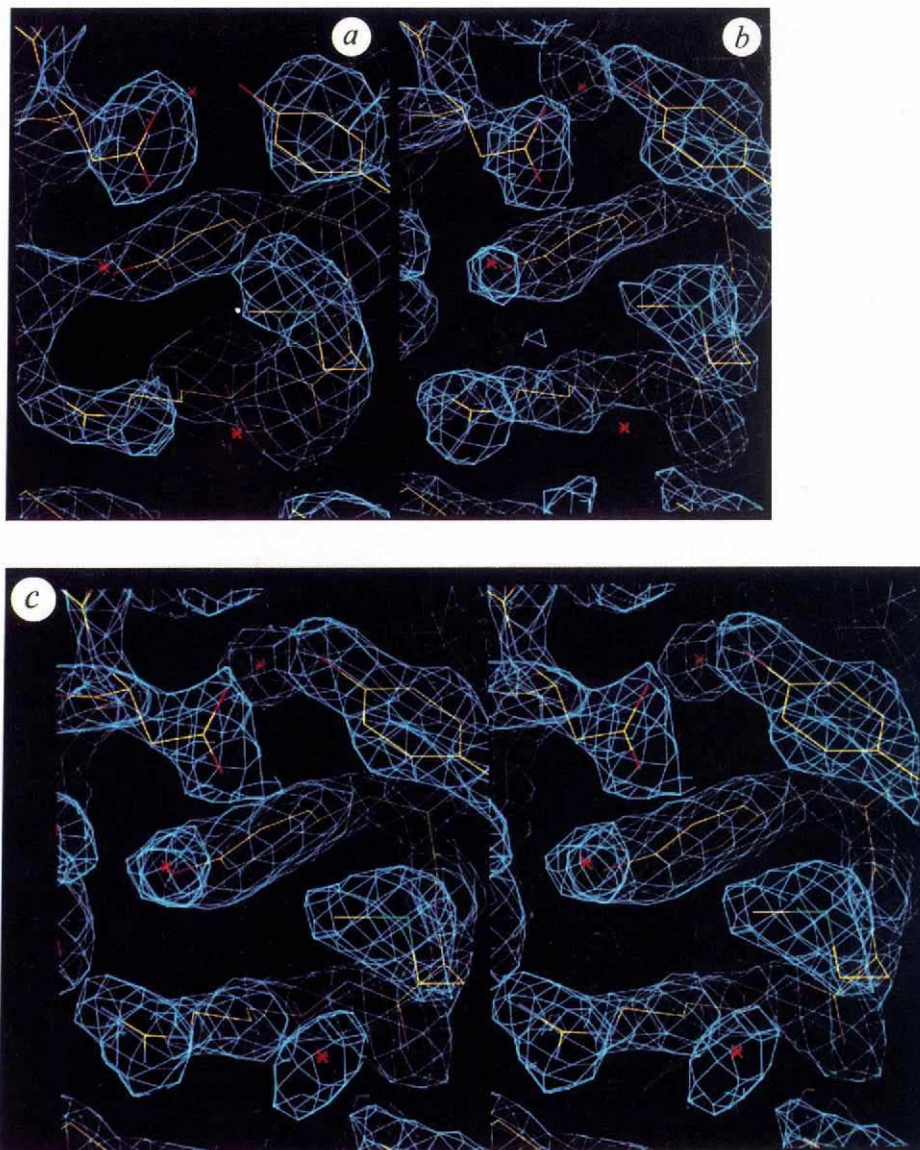
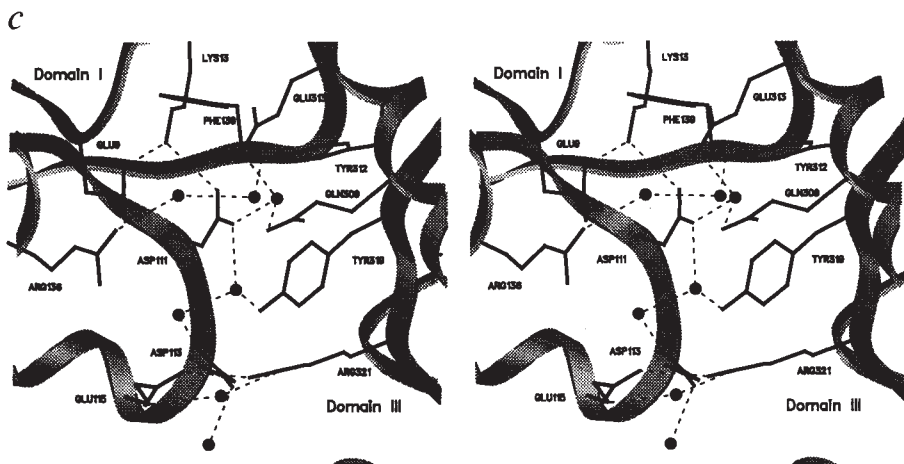
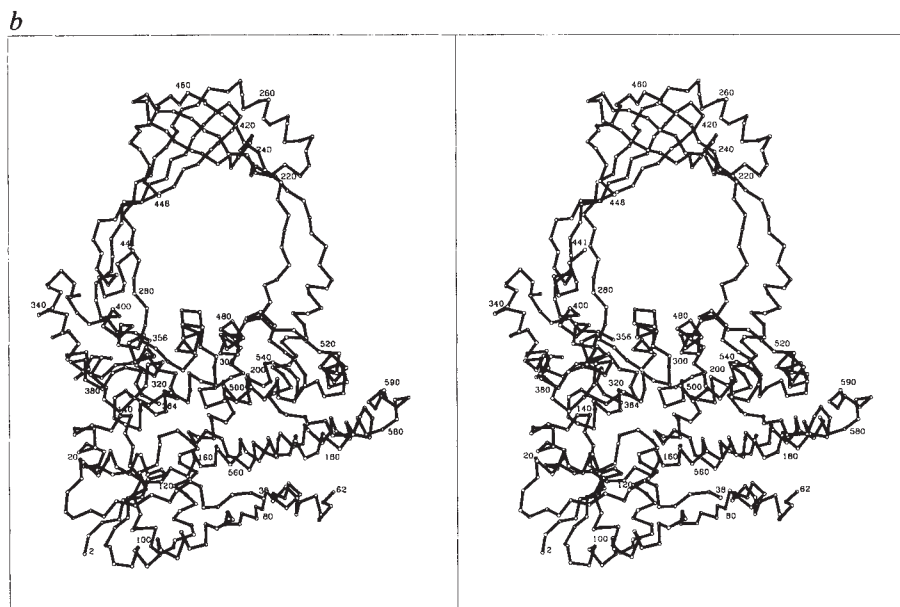
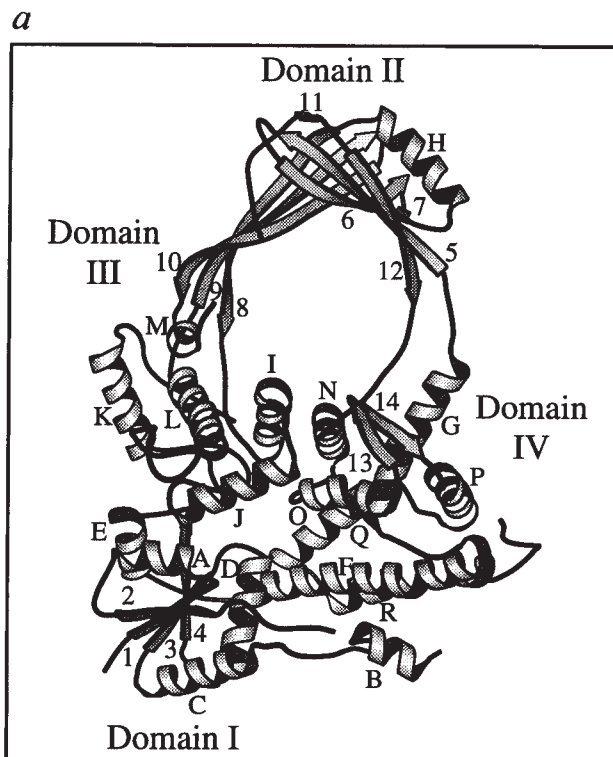


FIG. 1 Electron density maps. The electron density maps show the active site region of *E. coli* DNA topoisomerase I. In all cases the final, refined model is shown together with the electron density map. *a*, Solvent flattened multiple isomorphous replacement map calculated to 3.0 Å resolution. Contours shown correspond to a 2 σ level. This map was used to trace over 80% of the polypeptide chain. *b*, Simulated annealing omit map of the same region calculated to 2.2 Å resolution. Contours correspond to a 1 σ level. The omit map was generated by removing over 8% of the model inside a sphere centred on Tyr 319. Simulated annealing refinement of the partial model was used to remove bias²⁴. Sixteen overlapping omit maps were used to analyse the quality of the model throughout the entire structure of the 67K fragment. *c*, Stereo view of the same protein region. The final Sim-weighted (2Fo-Fc) map was calculated to 2.2 Å and corresponds to the model shown. Contours shown correspond to a 1.5 σ level.

FIG. 2 Overall architecture. *a*, Ribbon cartoon of the 67K fragment of *E. coli* DNA topoisomerase I generated by MOLSCRIPT⁶⁰. The four distinct domains are represented by Roman numerals, helices by letters A–R and β -strands by numbers 1–14. The N terminus begins just before strand 1. Four strands, 1–4, and four helices, A, C–E, make up domain I. A large loop, including helix B and a 25-amino-acid disordered region, is inserted between strand 2 and helix C. The carboxy-end of the sheets in this domain are involved in the active site region as well as the interface between domains I and III. Three acidic residues implicated in the coordination of Tyr 319 are located on the loop between strand 3 and helix D. The polypeptide chain then travels into domain IV, making up helices F and G. The proposed hinge region of the protein is located at the carboxy-end of helix G and the adjacent antiparallel strand after sheet 12. The polypeptide chain then travels into domain II comprising the antiparallel strands 5, 6 and 7 before turning into helix H. Once out of helix H, the chain makes up one of the antiparallel strands, 8, before heading into domain III. Domain III is made up of helices and some random coil. The polypeptide chain enters domain III from strand 8 into helix I. The chain proceeds to make up helices J and K before entering helix L. Helix J and the loop immediately thereafter, which contains the active site tyrosine, contain many other highly conserved amino acids in this topoisomerase subfamily. There is one small disordered region in the chain between helices K and L. Once the chain exits helix L, it proceeds back into domain II to complete anti-parallel strands 9 and 10. The chain briefly makes a turn back into domain III to complete the small helix M before re-entering domain II through a long, partially disordered random coil and strands 11 and 12. Once the chain has completed domain II, it travels back into domain IV through a coil between strand 12 in domain II and helix N in domain IV. The chain makes up helices N and O before completing the short antiparallel strands 13 and 14. From strand 14, the chain continues to complete helices P, Q and R with residues from all these helices helping to construct the large cleft located on this side of the protein. The chain ends some 20 amino acids beyond helix R and the carboxy-terminus is near helix P. There are extensive interactions involved in positioning domain III onto domains I and IV. The interactions between domains III and IV are made up of direct and water mediated hydrogen bonds. The major contacts occur between helix I in domain III and helix N in domain IV. The residues involved in direct hydrogen bonds are Q291/K484, K302/E479, V301/E479. Water mediated, hydrogen bonds occur between Q291/K487, S294/R515, K302/E479, S287/Y497. V301 lies on helix I making Van der Waals contact with both V483 and I501 on helix N. The interactions between domains I and III are more extensive. In addition to the direct and water-mediated hydrogen bonds, there is one salt bridge, E141/R369, and one water mediated salt bridge occurring between residues D113, E115 and R321. In the latter case, direct contacts are made between D113, E115 and R321; a water mediates the other contact between E115 and D321. Direct hydrogen bonds are made between E113/Y319, E111/Y312, S10/E313, N140/E366 and E313/T16. The through-water hydrogen bonds are made between E111/Y319, A314/K19, and the carbonyl oxygens of G315, I442 and Y20. *b*, Stereo view of the α backbone trace of the 67K fragment of *E. coli* DNA topoisomerase I. α positions are represented by balls and connected by solid sticks. Domain III contains the active site tyrosine at position 319. Residues forming the active site region of the enzyme are found between domains I, III and IV. *c*, The active site. Stereo representation of the active site tyrosine and neighbouring residues. Hydrogen bonding is represented by dashed lines between donors and acceptors. The polypeptide backbone is represented by a rectangular spline through the α position. Amino acids are labelled based on their three letter code and their position in the protein. The active site tyrosine is Tyr 319. SETOR and ancillary programs were used to generate all the figures unless otherwise noted⁶¹.



Electron density is clear for residues 2–38, 62–356, 364–441 and 448–590; discontinuity occurs at three places within surface loops of the protein molecule, the longest interruption being a 23-amino-acid stretch between residues 38–62. During refinement, the cross-validation (R_{free}) method was used as a measure of model bias²³. The R_{free} followed the same signature path as the working set R -factor (data not shown). As expected, the R_{free} dropped an additional 2% after the water molecules were added and the structure refined. In addition, simulated annealing omit maps were generated for sixteen distinct regions covering the entire protein²⁴. The resulting density in these regions agrees well with the final model. The R_{free} is 34.7%, and the final R_{work} is 22.4%, with r.m.s. deviations in bond lengths and angles of 0.014 Å and 1.75°, respectively.

Overall architecture

The 67K N-terminal domain of *E. coli* DNA topoisomerase I is made of four distinct domains which fold and contact each other to generate the shape of an elongated toroid with an overall dimension of 67 Å × 43 Å × 96 Å (Fig. 2). The large hole in the torus has an average side chain-to-side chain diameter of 27.5 Å, which is large enough to accommodate a piece of B-form DNA with no steric hindrance between the DNA phosphate backbone and protein side chains.

Domain I is made up of four parallel β -sheets, 1–4, sandwiched between four α -helices, two on top (A and E) and two below (C and D). This structure is similar to an α/β or Rossmann fold²⁵. Domain II is made up of three antiparallel β -strands numbered 5, 6 and 12 which cross a second set of three antiparallel β -strands numbered 8, 9 and 10. Helix H is located on one side and in between these packed sheets. The core of domain II is well ordered and filled with mostly aliphatic side chains. The curved surface for the upper half of the toroid results from the crossing of each set of twisted antiparallel β -strands in domain II at an angle close to 90°. Domain III contains the active-site tyrosine, Tyr 319, and is formed by four helices I, J, K and M packing around a fifth central helix, L. These helices appear to provide a strong base for the loops that contain many

of the highly conserved amino-acid residues, including Tyr 319 (Fig. 1; see also Figs 2 and 3). Domain IV is made up of mostly helical secondary structure. Four main helices F, G, Q and R make up this domain, upon which several other helices B, N, O and P pack. Antiparallel strands 14 and 15 are packed against helices G and N. Domain IV appears to provide a structural foundation for the protein.

The interface between domains I and III and between III and IV is extensive (Fig. 2). The interactions between domains III and IV involve both direct and water-mediated hydrogen bonds. The major contacts occur between helix I of domain III and helix N of domain IV. The interactions between I and III are more extensive. In addition to direct and water-mediated hydrogen bonds, there is one salt bridge and one water-mediated salt bridge. A feature of the interaction among these domains is the creation of a large cleft on one side of the protein and a small hole directly adjacent to the active site (Fig. 2). This hole runs parallel to the large hole of the toroid and may provide some of the space to accommodate single-stranded DNA in the vicinity of the active site residues. The C terminus of the protein is located adjacent to the large cleft. This side of the molecule is concave and proximal to the large disordered loops. It is plausible that the C-terminal domain of the intact *E. coli* enzyme, which is missing in the 67K fragment, is located on this side of the protein.

A striking feature of the 67K fragment is the path of the polypeptide chain as it travels through the various domains (Fig. 2). Domain I is made up of N-terminal residues. The polypeptide chain then travels through and makes up roughly one third of domain IV before proceeding into domain II. After completing over half of domain II, the chain travels into domain III. Once the chain completes domain III in its entirety, it heads back into domain II to finish the remaining half of this domain. Leaving domain II, the chain proceeds to complete domain IV. The interactions across the interface between domain III and domains I and IV are non-covalent, and the interface is devoid of chain entanglements. This architecture suggests domains II and III could move as a rigid body away from the rest of the protein. A fragment resistant to proteolysis has been found that supports this idea. Cleavage by chymotrypsin after Phe 214 and around amino acid 470 yields a stable 30K fragment containing domains II and III. The two cleavage sites are opposite each other in the

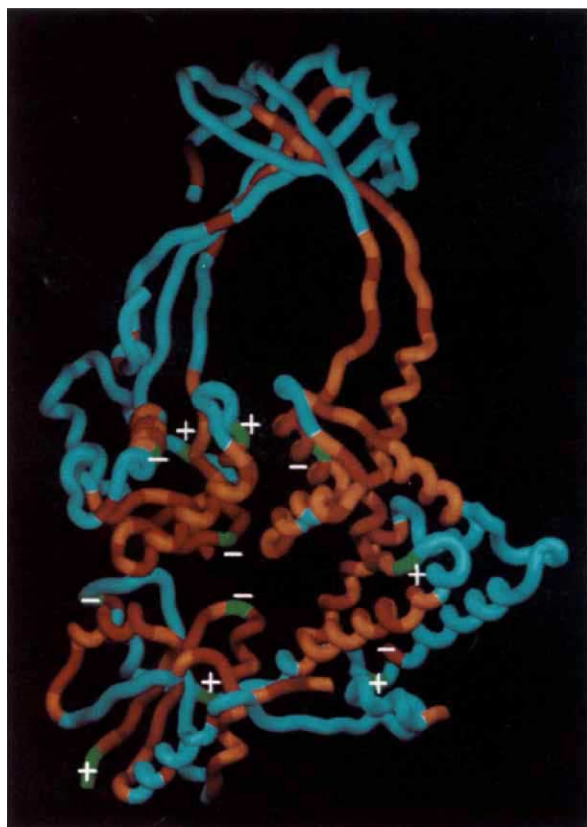


FIG. 3 Conserved sequence motifs. Worm diagram generated by a spline through the $C\alpha$ backbone position with colour representation of the strictly conserved residues (red) and regions of high homology (yellow) on a blue background. Regions of high homology include the active site region, the cleft adjacent to the active site and the helix and strands comprising the proposed hinge region between domains II and IV. The sequence comparison was done between five members of this subfamily of DNA topoisomerases, including *E. coli* DNA topoisomerase I; *E. coli* DNA topoisomerase II (ref. 9); *S. acidocaldarius* reverse gyrase¹³; *S. cerevisiae* DNA topoisomerase III (ref. 10); *Synechococcus* sp. *topA* gene product (GenBank X72391); and *Bacillus firmus* *topA* gene product⁶². The location of the regions of high sequence similarity and the novel structure of the *E. coli* enzyme leads to the conclusion that other members of the subfamily are most likely to contain domains similar to those found in the 67K fragment of the *E. coli* DNA topoisomerase I. The location of the insertions of tetrapeptides into the structure are marked in green. The – and + signs indicate a total loss of enzymatic activity or a partial retention of activity, respectively. It is interesting to note that several of the insertions which cause loss of activity are located in the active site region of the protein or inside key secondary structural regions. Most of the insertions which do not remove all enzymatic activity are located in loops between secondary structural elements and are away from the active site region of the protein.

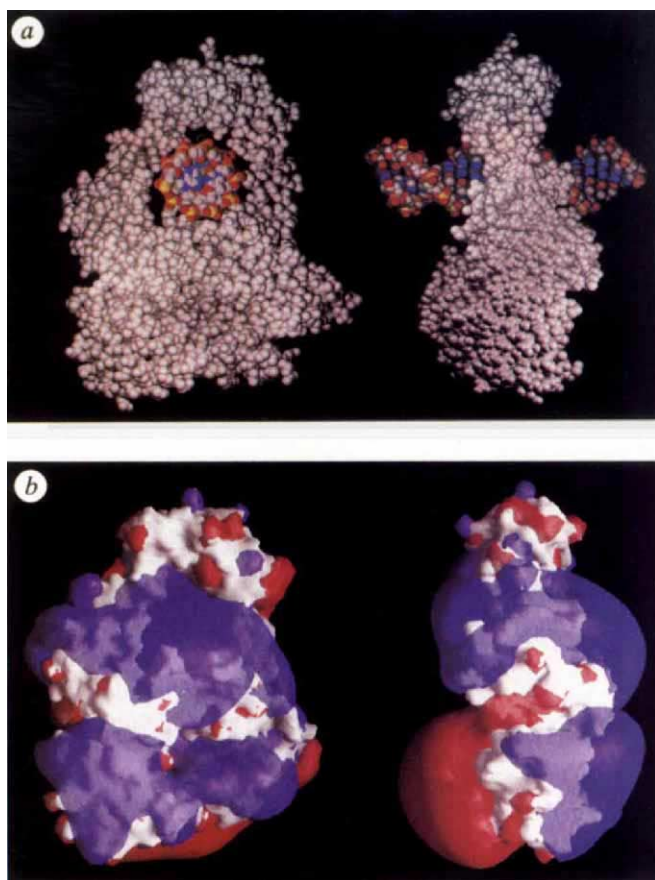
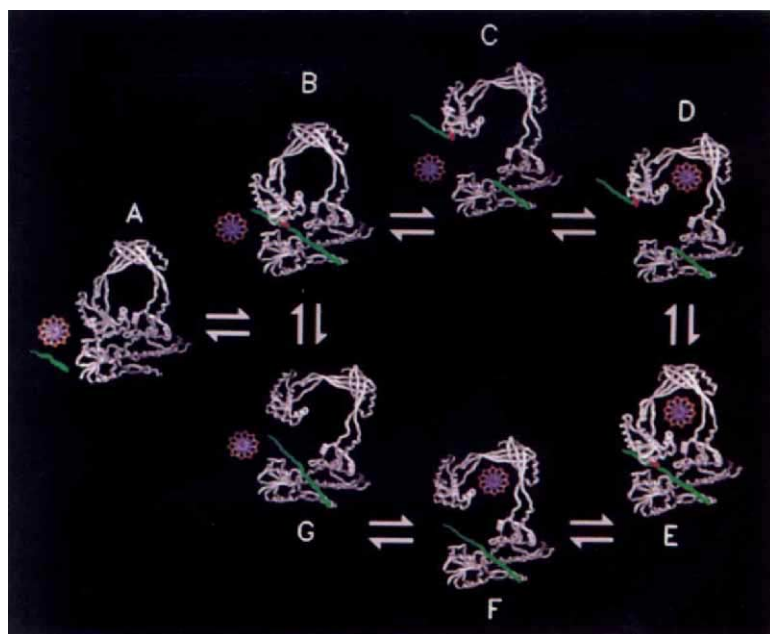


FIG. 4 Possible involvement of an intermediate in which a B-form DNA is enclosed by the toroidal enzyme. *a*, Possible interaction between a B-form DNA in the large hole of *E. coli* DNA topoisomerase I. Molecules are represented by van der Waals surfaces. The two views are orthogonal to each other. The DNA axis is shown perpendicular to the longest dimension of the protein and parallel to the thinnest dimension. As mentioned in the text, DNA may be located within the torus during the formation of a transient complex. The complex cannot be stable, otherwise these intermediates would become trapped in a thermodynamic valley and inhibit the reaction. *b*, Electrostatic potential map generated with the program DELPHI³⁶ with contours of negative (red) and positive (blue) potential at $\pm 2.5 k_B T/e$ ($k_B T$ being the product of the Boltzmann constant and the temperature, e is the charge of an electron). GRASP was used to display the molecular model and potential map³⁷. The electrostatic calculations used a dielectric constant of 80 for water and 2 for protein. The contours shown are for a potential map calculated in a nonionic solvent (0 mM salt). Potential maps calculated with 145 mM salt show similar features. The map was calculated using fully charged lysines, arginines, aspartic and glutamic acids. A half-charge was assigned to the histidine residues. Eighteen lysines and arginines line the hole compared to only nine aspartic and glutamic acid residues. Only two histidines are in the proximity of the hole.

FIG. 5 Proposed steps in the strand passage reaction. The possible intermediate steps are illustrated using the strand passage reaction associated with catenation and decatenation of double-stranded rings where one ring has a pre-existing nick or gap. The nick or gap provides the single-stranded region within which the enzyme cleaves during catalysis. In this reaction, the single-stranded region is cleaved and the duplex portion of a different DNA molecule is passed through the break. This reaction would require an intermediate of a duplex DNA inside the torus. The complex formed is expected to be transient and unstable. In the following scheme, the passing strand (Pex or Pin) is represented by the coloured duplex DNA molecule and the gated strand (Gc or Go) is represented by the green tube. A, (Pex + Ec + Gc). In this state, both the passing (Pex) and gated (Gc) strands are not yet associated with the protein. The enzyme and the gated strand are closed, as denoted by the subscript (c), and the passing element is external to the torus, as denoted by the subscript (ex). It should be emphasized that no directionality for this mechanism is known and that all intermediates could proceed in either direction as indicated by the double arrows between each step. Also, no one intermediate can be thermodynamically very stable, otherwise the molecules would be trapped in a local energy minimum. B, (Pex + EcGc). The passing element is still external to the protein and the gated strand is now interacting noncovalently with domains I, III and IV of the protein. The gated strand is still closed or continuous and the protein is closed. The enzyme would have to open slightly in order to allow the gated strand access to this region of the protein. C, (Pex + EoGc). The passing element is external to the protein and both the gated strand and protein are now open. Tyr 319 is shown covalently attached to the 5'-phosphoryl end of the gated or broken strand (red dot). The 3'-OH end of the gated strand is shown interacting noncovalently with domain I and the cleft involving domain IV. D, (Pin + EoGc). The protein and gated strand are in the same state as in step C, but Pex has passed through the break



or gate created by separating the broken ends of the DNA. The passing element is now internal, Pin, to the torus and has passed through the break. E, (Pin + EcGc). The passing element is transiently trapped within the torus while the enzyme closes to reseal the gated strand. F, (Pin + EoGc). The enzyme and gated strand are interacting noncovalently, but the enzyme is in the open form; the passing element is still internal to the torus but can now exit from the hole. G, (Pex + EoGc). The enzyme remains open and interacts noncovalently with the gated strand; the passing element has come out of the torus.

TABLE 1 X-ray structure determination

Spacegroup P2 ₁ 2 ₁ 2 ₁	Cell $a = 63.19$ $b = 77.97$ $c = 138.90$				
Topo67	Native I	PheHg	Iodoacetamide/ AuCN ₂	PTCN ₄	Native II
Detector/ λ	Xentronics/Siemens/CuK α /1.54 Å				
Max. resolution (Å)	2.61	2.63	2.71	3.23	2.11
Unique total reflections	21,717/115,565	21,094/118,637	18,584/106,554	10,676/64,255	38,618/344,848
Data coverage (%)	99.6	97.5	93.9	96.0	96.0
R_{sym} (%) [*]	6.5	6.4	7.9	9.9	6.5
MFID [†]		14.6	25.8	14.5	
MIR analysis (50–3 Å)					
Max. resolution (Å)		3.0	3.0	4.5	
Number of sites		4	10	12	
Phasing power [‡]		1.1	1.2	0.9	
R_c		0.83	0.79	0.86	
Mean FOM	0.497				
Refinement (SSRL IP data)					
Resolution (Å)		5–2.2			
R -factor [¶]		0.224			
R_{free} [#]		0.347			
R_{overall} ^{**}		0.235			
Reflections (total)		32,009			
Working set		28,794			
Test set		3,215			
Number of total atoms		4,679			
Protein/water atoms		4,412/267			
R.m.s. bond ^{††}		0.014			
R.m.s. angle ^{††}		1.75			
Average B factor ^{‡‡}		26.8			

Data collection. Expression, purification and crystallization of the 67K fragment of *E. coli* DNA topoisomerase I has been reported previously¹⁹. Typical crystals are 150 × 250 × 400 microns. Diffraction data were collected using a Xentronics/Siemens area detector^{47,48} mounted on a Rigaku RU200 generator (0.1 mm focusing cup). The X-ray beam was focused with Franks mirrors⁴⁹ on the crystal position to maximize sample diffraction. The data were processed using XDS and XSCALE^{50,51}. Data from native and isomorphous derivative crystals were initially collected at −4 °C. Diffraction from crystals exposed to X-rays at room temperature decay within the first 5 min of exposure. The decay can be reduced by cooling to −4 °C, although high-resolution diffraction decay to 3.2 Å within 20 min of exposure where it then remains stable for 2–3 days. MIR phases derived from these data sets did not produce interpretable electron density maps. To overcome these problems, crystals were cryocooled to −168 °C (ref. 52) with a Siemens LTI cryocooler and native and derivative data sets were re-collected. The high-resolution native data set used for refinement was collected at beam line X8C at Stanford Synchrotron Radiation Laboratory at 1.08 Å wavelength using a MAR scanner and a modified Siemens LTI cryocooler. Stanford Synchrotron Radiation Laboratory data were reduced using DENZO (Z. Otwinowski, personal communication). All subsequent scaling and processing was done using the CCP4 suite⁵³ unless otherwise noted. **Heavy atom derivatives.** Crystals were transferred from 2.15 M to 2.8 M (NH₄)₂SO₄ for stabilization and derivatization. The pH of this unbuffered solution was 5.35. Several mercurial compounds were tried; all but one, phenylmercuric acetate, produced large non-isomorphous differences and poor phases. Crystals were prepared by transferring to a 0.5 mM phenylmercuric acetate solution for 12 h at room temperature followed by a 24 h back soak at 4 °C. One site was located using Patterson methods. Single isomorphous replacement phases from this one-site derivative were used to solve the other derivatives. The gold derivative was initially prepared by transferring the crystal in 5 steps to 5 mM gold (I) potassium cyanide (KAuCN₂) for 6 h at room temperature. Initial analysis showed a peak at the same position as the mercury with very non-isomorphous results. This site was assumed to be the only cysteine (Cys 401) in the protein. In subsequent gold derivatizations, the cysteine was first blocked by acetylation by transferring the crystal to 10 mM iodoacetamide, pH 5.35, for 2 h at room temperature to prevent acetylation of other sites and to ensure isomorphous conditions. Following acetylation, crystals were transferred in 5 steps to 5 mM KAuCN₂ for 6 h at room temperature. This yielded isomorphous data which revealed two major gold sites. The potassium tetracyanoplatinate (II) K₂(PtCN₄) derivative was prepared by slowly transferring crystals in 10 steps to 10 mM K₂(PtCN₄) while keeping them on ice to prevent fissuring. These crystals were stored at 4 °C for 5 h before freezing. **Cryocooling.** All crystals were transferred on ice to a 20% sucrose, 2.8 M (NH₄)₂SO₄ solution in 10 steps before cryocooling at −168 °C. Among the many cryo-buffers tried, only sucrose produced isomorphous freezing conditions for all derivative and native crystals. All other cryoprotectants tried produced large non-isomorphous changes with the mercury derivative. Crystals were mounted using a rayon loop⁵⁴. One crystal was used for each native and derivative data set. **Phasing, model building and refinement.** Heavy-atom refinement and phasing was done using MLPHARE⁵⁵. An initial MIR electron density map was calculated to 3.0 Å using the major sites from the mercury and gold derivatives. Secondary structural elements and some side chains were readily apparent. Solvent flattening^{56,57} to 3.0 Å was done to improve the electron density map. Given the excellent quality of the phases to 4.5 Å, the mask was only calculated to this resolution. A non-continuous polyaniline chain was traced into the density with the program O, using the Bones and Lego options²⁰. Side chains were added where apparent and initial positional refinement was carried out using PROLSQ²¹. Combined model and MIR phases were subsequently used to calculate 2Fo – Fc and Fo – Fc maps to further interpret and modify the model. At this point, the platinum derivative and all the minor sites were added to the MIR phasing calculations. Although the platinum phasing statistics are poor, its inclusion helped to improve the connectivity of the MIR electron density map. Subsequent refinement with XPLOR was done against high resolution data collected at Stanford Synchrotron Radiation Laboratory²². Two rounds of simulated annealing, positional refinement and manual model rebuilding to 2.5 Å were done. Sim weighted⁵⁸ 2Fo – Fc and Fo – Fc maps were used to check the model. Grouped B factors were refined; the R -factor at this stage was 25.7%. The model stereochemistry and geometry was analysed using PROCHECK⁵⁹ and corrected. The current model has no outliers in the Ramachandran plot. Another round of simulated annealing was done using data between 5 and 2.2 Å. Wherever necessary the model was corrected and over 200 water molecules were added. Final refinement gave an R -factor of 22.4% and an R_{free} of 34.7%. The final model includes 4,412 protein atoms and 267 water atoms. The coordinates will be deposited in the Brookhaven Protein Data Bank.

^{*} $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I = observed intensity, and $\langle I \rangle$ = average intensity obtained from multiple measurements.

[†] MFID (mean fractional isomorphous difference) = $\sum |F_{\text{ph}}| - |F_{\text{p}}| / \sum |F_{\text{p}}|$, where $|F_{\text{p}}|$ = protein structure factor amplitude and $|F_{\text{ph}}|$ = heavy-atom derivative structure factor amplitude.

[‡] Phasing power = root-mean-square ($|F_{\text{h}}|/E$), where $|F_{\text{h}}|$ = heavy-atom structure factor amplitude and E = residual lack of closure error.

[§] $R_c = \sum |F_{\text{h}}(\text{obs})| - |F_{\text{h}}(\text{calc})| / \sum |F_{\text{h}}(\text{obs})|$ for centric reflections where $|F_{\text{h}}(\text{obs})|$ = observed heavy atom structure factor amplitudes, and $|F_{\text{h}}(\text{calc})|$ = calculated heavy-atom structure factor amplitude.

^{||} FOM, figure of merit.

[¶] R -factor, R factor based on 90% of the native data used in refinement.

[#] R_{free} , R factor based on 10% of the native data withheld for the cross-validation test.

^{**} R_{overall} , R factor based on all data including data withheld for the R_{free} test.

^{††} R.m.s. bond angles and r.m.s. bond lengths are the respective r.m.s. deviations from the ideal values.

^{‡‡} Average B factor, average value of B factors for all atoms in structure.

two strands joining domain IV to domains II and III (C.D.L. and A.M., unpublished results). The rigid body movement of these domains could provide an adequate opening into the large hole for either single- or double-stranded DNA and immediately implies the existence of both an open and a closed form of the enzyme.

Active site

The catalytic tyrosine, Tyr 319, is located on domain III at the interface made by domains I, III and IV. This interface is located next to the large cleft and small hole in the main body of the protein (Figs 1 and 2). Tyrosine 319 and its neighbouring amino acids are involved in some of the major interactions among these domains; Tyr 319 interacts with several highly conserved amino acids, including Glu 9, Lys 13, Asp 111, Asp 113, Glu 115, Gln 309, Glu 313, Tyr 312 and Arg 321, to form an extensive network of hydrogen bonds. Several water molecules and a water-mediated salt bridge are also involved in this network (Fig. 2). Tyrosine 319 is hydrogen-bonded through one water molecule to Asp 111 and through two water molecules to Asp 113 and Glu 115. Direct contact is made between Asp 113, Glu 115 and Arg 321, with a water mediating the salt bridge between Glu 115 and Arg 321; Tyr 312 is hydrogen-bonded to the backbone carbonyl oxygen of Phe 139 and through two water molecules to Thr 110 and Arg 136; Lys 13 forms a hydrogen bond to Glu 9 and Asp 111, providing additional contacts between domain I and the residues directly involved in Tyr 319 coordination. A second shell of conserved amino acids, including Gly 32, His 33, Arg 168, Gln 309, Ala 308, Leu 311, Glu 312, Ile 317, Thr 318, Thr 322, Asp 323, Ser 324, His 365, Ala 367, Ile 368, Tyr 391, Thr 496, Ala 498 and Glu 547, surrounds the core of residues clustering around the active site tyrosine. An interesting aspect of the active site is the location of the three acidic residues Asp 111, Asp 113 and Glu 115: the spatial similarity of these residues to the acidic residues that coordinate two divalent cations in the exonuclease catalytic site of Klenow fragment is striking²⁶. Magnesium has been implicated in the cleavage reaction of single-stranded DNA²⁷ but no bound metal ion was identified in the structure of the 67K fragment.

Highly conserved regions and mutations

Highly conserved regions of the protein sequences amongst members of the *E. coli* DNA topoisomerase I subfamily map to parts of the 67K structure that appear to be important for function (Fig. 3). In particular, residues located in the vicinity of the active-site tyrosine, the large cleft next to the active site, and the region between domains II and IV are highly conserved. Few highly conserved residues are found in domain II despite its clear structural role. Although the interior of the toroid exhibits rather moderate sequence similarity, its overall charge character and distribution appears to be similar within the subfamily.

The structural domains of the protein have been probed by inserting tetrapeptides into the protein and testing for relaxation activity and single-stranded DNA cleavage²⁸. Thirteen such mutations map to the 67K fragment. Of those, one is within a disordered region of the structure. The other twelve are shown in Fig. 3, marked with either a plus or minus sign to indicate whether or not the enzyme retained partial activity. Many of the mutations that abolish activity map to the active-site region, whereas mutations that preserve catalytic activity tend to be located in loops between secondary structural elements. Not surprisingly, the mutations that cause inactivation tend to coincide with regions of high sequence similarity to other members of the subfamily.

Mechanistic implications

E. coli DNA topoisomerase I can have several conformations.

The structure of the 67K fragment shows that the active-site tyrosine is buried between two domains in the centre of the protein, thus making the active-site tyrosine inaccessible to

single-stranded DNA. Both the intact protein and the 67K fragment are capable of cleaving single-stranded DNA, which strongly suggests that the protein undergoes a conformational change to expose the active-site tyrosine. The novel architecture of the enzyme suggests that domains II and III can be lifted off domains I and IV as a rigid body, using the strands joining domains II and IV as a hinge. The putative hinge corresponds to the region where several proteases cleave to separate domains II and III from the rest of the protein. This kind of domain movement is particularly attractive from a mechanistic point of view, as the active-site tyrosine is located in domain III at a strategic position where domains I, III and IV converge.

The cleavage step of the reaction involves a covalent bond between Tyr 319 and the 5'-phosphoryl end of the broken strand. The 3'-OH end of the broken strand is not covalently associated with the protein. As a convention, the covalently attached end of the cleaved strand is referred to as the plus end and the non-covalently attached end the minus end. If the cleavage of single-stranded DNA is coupled to the movement of domains II and III away from the rest of the protein, then the conformational change would create a DNA gate as long as the minus end remains associated with the rest of the protein as the domain III-linked plus end is lifted away. The gate could be wide enough to accommodate the passage of either a single- or a double-stranded DNA through the break. Under physiological conditions, such a conformational change is presumably triggered solely by protein-DNA interactions as no cofactor, such as ATP, is needed for enzymatic activity. This contrasts with the mechanism of the type-II enzymes where ATP-mediated conformational changes occur during DNA manipulations^{29,30}.

Enzyme-DNA interactions: the cleaved strand. As already discussed, cleavage of the DNA substrate is likely to be followed by a widening of the distance between the ends of the broken strand to allow strand passage. The minus end is likely to interact non-covalently with either domain I or IV and remain associated with the protein. Analysis of the nucleotide sequences around the cleavage sites shows a strong preference for a cytosine four bases away from the 3'-OH end of the minus end. This base preference suggests the existence of specific interactions between the protein and the minus end. Domain I and the large cleft seem to be particularly well posed for interacting with the minus end. In support of this role for domain I is its similarity to the α/β or Rossmann fold found in many proteins. In many instances the Rossmann fold has been associated with nucleotide or dinucleotide binding. Domain I was fit to similar domains in four α/β proteins known to bind nucleotides; lactate dehydrogenase³¹, alcohol dehydrogenase³², *p*-hydroxybenzoate³³ and glutathione reductase³⁴. The fit was done against the β -strands and gave an r.m.s. difference of ~ 2.2 Å between α -carbon atoms in respective domains (data not shown), which is comparable to the r.m.s. difference found when the proteins are compared to each other. Hence, it is tempting to associate domain I, which resembles a nucleotide-binding domain, with the region where the tip of the minus end is bound.

The cleavage of gapped DNA molecules by the *E. coli* enzyme occurs at or near the junction between the single- and double-stranded regions of the substrate³⁵. The double-stranded region corresponds to the minus end. The gapped molecule is likely to orient itself with the minus end close to domains I and IV and towards the C terminus of the 67K fragment. The cleft adjacent to the active-site tyrosine in the crystal structure appears to be an attractive choice for the binding of double-stranded DNA. Electrostatic calculations^{36,37} show that the large cleft is shrouded in positive potential suitable for DNA binding. In the intact *E. coli* enzyme, domain IV extends into the C-terminal domain containing three tetracysteine motifs, which is likely to be involved in binding of DNA³⁸.

Enzyme-DNA interactions: the central hole in the toroidal protein. A striking feature of the 67K protein structure is the torus shape with a 27.5 Å (side chain to side chain) average diameter

hole. The toroidal architecture is clearly needed to provide a hinge so that domain III can be moved away from the main body of the protein, and thus opening a gate to allow strand passage. The torus may also play a role in accommodating either single- or double-stranded DNA, after or before strand passage (Fig. 4a). The interactions between DNA and the protein residues lining the hole are necessarily nonspecific and weak, otherwise a stable DNA-protein complex would form, trapping the reaction intermediates. The size of the hole can comfortably accommodate either single- or double-stranded DNA without steric hindrance. Modelling of DNA through the hole is supported by electrostatic calculations^{36,37} indicating the positively charged nature of the hole. Calculations at two salt concentrations show the same overall features: the hole is immersed in positively charged potential that could counter the strong negative potential of the DNA phosphate backbone (Fig. 4b).

The structure of the 67K fragment does not contain any known DNA-binding motif that would suggest binding to the major or minor grooves of DNA. In this sense, it represents a novel protein fold and DNA-binding motif. The other protein known to surround DNA when bound is the processive element of DNA polymerase III holoenzyme³⁹. The protein interacts nonspecifically with DNA and glides along the DNA helix, providing the processive element. The electrostatic field of the holes of both proteins are very similar. The 67K fragment has a much smaller hole than the processive element (27.5 Å compared to 35 Å, side chain to side chain), which may serve to keep *E. coli* DNA topoisomerase from forming a stable protein-DNA complex. It may also serve to give the *E. coli* enzyme the ability to sense secondary structure within the DNA.

Domain II of the *E. coli* enzyme fragment shows some overall topological similarity to the TATA-binding factor (TBP or TFIID)⁴⁰⁻⁴². Unlike the *E. coli* enzyme, TBP is involved in sequence-specific DNA binding. Domain II of the *E. coli* enzyme can be placed directly onto TBP to show an identical radius of curvature. This curvature, which presumably allows the proteins to partially surround DNA, is due to the twisted β sheet structure of TBP and domain II of the *E. coli* enzyme. The similarity to TBP seems at present to extend only to the shape of domain II.

DNA passage by *E. coli* DNA topoisomerase I. The structure of *E. coli* DNA topoisomerase I clearly suggests at least two conformations, a closed form (E_c), in which domain III is intimately associated with the rest of the protein, and an open form (E_o) where domain III has moved away from the base of the protein, thus exposing the catalytic tyrosine and opening the large torus. During the cleavage step, the two ends of the broken strand are separated to form an opening or gate that allows strand passage. This mechanism has been described as the enzyme bridging mechanism (refs 14, 35, 43-45; for a review, see ref. 46). That is, the enzyme serves as a bridge to create a gap in a DNA strand by mediating covalent and noncovalent interactions with both ends of the broken strand. The strand to be cleaved or gated then also has two forms: an intact or closed form (G_c) and a broken or open form (G_o). To account for the strand passage reaction, the enzyme and the DNA gated strand must form a minimal of three complexes that are distinguishable by the open or closed states of the enzyme and the DNA gated strand: a complex E_cG_c in which the enzyme is in the closed form and the DNA gated strand is unbroken, a second complex E_oG_o in which both the enzyme and the gated strand are open, with the plus end of the broken gated strand covalently attached to domain III of the enzyme and the minus end non-covalently bound to the domains I and/or IV of the enzyme, and a third complex E_oG_c , in which an intact DNA gated strand is interacting non-covalently with the enzyme in its open form. In the E_oG_o state, a second DNA strand or duplex can pass from the exterior of the enzyme through the DNA gate to its interior, or vice versa. In the E_oG_c state, a second DNA strand or duplex can pass from the exterior of the enzyme to its interior, or vice versa

without going through the DNA gate. Both types of reactions are necessary for an enzyme molecule to catalyse multiple cycles of reactions without complete dissociation of the gated strand from the enzyme in each cycle.

Figure 5 illustrates one plausible scheme for a catenation reaction in which a duplex DNA ring is passed through a nicked or gapped DNA ring; the same scheme can be applied to other topological transformations of DNA rings by the enzyme. Only a short single-stranded segment is shown for the gate strand and the double-stranded DNA segment being passed through the gate is viewed from its end; P_{ex} and P_{in} denote the passing DNA being external or internal to the toroidal enzyme, respectively. In all of the steps illustrated, the interactions between the protein and the passing strand must be weak, otherwise a stable complex would be formed, thus trapping a reaction intermediate in a thermodynamic valley. Because the reaction proceeds in the absence of any external energy source, all of the steps must be driven by protein-DNA and protein-protein interactions, and the directionality of the overall reaction must be determined by the free energies of the DNA reactants and products in the presence of catalytic amounts of the enzyme. The C-terminal domain of the intact enzyme, which is missing in the 67K structure, must be involved at some point in this mechanism, as it is required for the strand passage reaction. The relaxation of a duplex DNA molecule would require the separation of the two strands into distinct single-stranded regions to facilitate the strand passage reaction. It is known that *E. coli* DNA topoisomerase I requires single-stranded regions in duplex DNA for the relaxation reaction.

The proposed model for strand passage based on biochemical data for the intact *E. coli* DNA topoisomerase I and fragments derived from it, and the three-dimensional structural data of the 67K fragment, provides a molecular view of the major aspects of the many protein-DNA interactions that must occur for the enzyme to transform DNA topologically. It supplies a testable model consistent with the available mechanistic and structural data and provides the first structural hint of how this subfamily of type-I DNA topoisomerases perform complicated topological transformations of DNA. □

Received 15 October; accepted 18 November 1993.

- Gellert, M. A. *Rev. Biochem.* **50**, 879-910 (1981).
- Wang, J. C. A. *Rev. Biochem.* **54**, 665-697 (1985).
- Wang, J. C. *J. Biol. Chem.* **266**, 6659-6662 (1991).
- Proc. Int. Symp. on DNA Topoisomerases in Chemotherapy* Nagoya Japan (eds Angho, T., Ikeda, H. & Oguro, M.) (CRC, Florida, 1993).
- Liu, L. F. A. *Rev. Biochem.* **58**, 351-375 (1989).
- Drlaca, K. & Franco, R. *J. Biochemistry* **27**, 2253-2259 (1988).
- Wang, J. C. *J. molec. Biol.* **55**, 523-533 (1971).
- Dean, F. et al. *Cold Spring Harbor Symp. quant. Biol.* **47**, 769-777 (1982).
- DiGate, R. J. & Marians, K. J. *J. Biol. Chem.* **264**, 17924-17930 (1989).
- Wallis, J. W., Chrebet, G., Brodsky, G., Roffe, M. & Rothstein, R. *Cell* **58**, 409-419 (1989).
- Kim, R. A. & Wang, J. C. *J. Biol. Chem.* **267**, 17178-17185 (1992).
- Kikuchi, A. & Asai, K. *Nature* **309**, 677-681 (1984).
- Confalonieri, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4753-4757 (1993).
- Wang, J. C. in *Nucleases* (eds Roberts, R. & Linn, S.) 41-57 (Cold Spring Harbor Laboratory, New York, 1982).
- Kornberg, A. & Baker, T. *DNA Replication* (Freeman, New York, 1992).
- Tse-Dinh, Y.-C. & Beran-Steed, R. K. *J. Biol. Chem.* **263**, 15857-15859 (1988).
- Lynn, R. & Wang, J. C. *Prot. Struct. Func. Genet.* **6**, 231-239 (1989).
- Tse-Dinh, Y.-C. *J. Biol. Chem.* **266**, 14217-14320 (1991).
- Lima, C. D., Wang, J. C. & Mondragón, A. *J. molec. Biol.* **232**, 1213-1216 (1993).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta. crystallogr.* **A47**, 110-119 (1991).
- Hendrickson, W. A. & Konner, J. H. in *Biomolecular Structural, Function, Conformation and Evolution* Vol. I (ed. Srinivisan, R.) 43-57 (Pergamon, Oxford, 1980).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Brünger, A. T. *Nature* **355**, 472-474 (1992).
- Hodel, A., Kim, S.-H. & Brünger, A. T. *Acta crystallogr.* **A48**, 851-858 (1992).
- Brändén, C.-I. *Quart. Rev. Biophys.* **13**, 317-338 (1980).
- Beese, L. S. & Steitz, T. A. *EMBO J.* **10**, 25-33 (1991).
- Dominico, P. L. & Tse-Dinh, Y.-C. *J. inorg. Biochem.* **42**, 87-96 (1991).
- Zumstein, L. & Wang, J. C. *J. molec. Biol.* **191**, 333-340 (1986).
- Lindsay, J. & Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10485-10489 (1991).
- Roca, J. & Wang, J. C. *Cell* **71**, 833-840 (1992).
- Grau, U. M., Tromer, W. E. & Rossmann, M. G. *J. molec. Biol.* **151**, 289-307 (1981).
- Eklund, H., Samana, J.-P. & Jones, T. A. *Biochemistry* **23**, 5982-5996 (1984).
- Van der Laan, J. M. et al. *Biochemistry* **28**, 7199-7205 (1989).
- Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **195**, 701-729 (1987).
- Kirkgaard, K., Pflugfelder, G. & Wang, J. C. *Cold Spring Harbor Symp. quant. Biol.* **49**, 411-419 (1984).

36. Gilson, M. K., Sharp, K. A. & Honig, B. H. *J. comp. Chem.* **9**, 327–335 (1987).
37. Nicholls, A., Sharp, K. A. & Honig, B. H. *Prot. Struct. Funct. Genet.* **11**, 281–286 (1991).
38. Beran-Steed, R. K. & Tse-Dinh, Y.-C. *Prot. Struct. Funct. Genet.* **6**, 249–258 (1989).
39. Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. *Cell* **69**, 425–437 (1992).
40. Nikolov, D. B. et al. *Nature* **360**, 40–46 (1992).
41. Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **356**, 520–527 (1993).
42. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–520 (1993).
43. Brown, P. O. & Cozzarelli, N. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 843–847 (1981).
44. Tse, Y.-C. & Wang, J. C. *Cell* **22**, 269–277 (1980).
45. Kirkegaard, K. & Wang, J. C. *J. molec. Biol.* **185**, 625–637 (1985).
46. Champoux, J. J. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 217–242 (Cold Spring Harbor Laboratory Press, New York, 1990).
47. Durbin, R. M. et al. *Science* **232**, 1127–1132 (1986).
48. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Cryst.* **20**, 235–242 (1987).
49. Harrison, S. C. *J. appl. Cryst.* **1**, 84–90 (1968).
50. Kabsch, W. *J. appl. Cryst.* **21**, 67–71 (1988).
51. Kabsch, W. *J. appl. Cryst.* **21**, 916–924 (1988).
52. Hope, H. *Acta. crystallogr.* **B44**, 22–26 (1988).
53. CCP4 SERC (UK) Collaborative Computing Project 4 (Daresbury Laboratory, Warrington WA4 4AD, 1979).
54. Teng, T.-Y. *J. appl. Cryst.* **23**, 387–391 (1990).
55. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering* 80–86 (Daresbury Laboratory, UK (1991)).
56. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
57. Leslie, A. G. W. *Acta. crystallogr.* **A43**, 134–135 (1987).
58. Sim, G. A. *Acta. crystallogr.* **13**, 511–512 (1960).
59. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Cryst.* **26**, 283–291 (1993).
60. Kraulis, P. J. *J. appl. Cryst.* **24**, 946–950 (1991).
61. Evans, S. V. *J. molec. Graph.* **11**, 134–138 (1993).
62. Ivey, D. M., Cheng, J. & Krulwich, T. A. *Nucleic Acids Res.* **20**, 4928 (1992).

ACKNOWLEDGEMENTS. We thank A. Sharma and N. Lue for help with data collection and for their comments; the staff at BNLS, SSRL, the Photon Factory and CHESS for their help; P. Bjorkman, D. Linzer and J. Widom for discussion; L. Beese, J. Kuriyan and S. Burley for coordinates; and our friends and colleagues for their support. This work was supported by grants from the NIH. Final refined coordinates will be deposited in the Brookhaven Protein Data Bank; in the interim they can be requested by E-mail to a.mondragon@nwu.edu.

LETTERS TO NATURE

Superconductivity at 23 K in yttrium palladium boride carbide

R. J. Cava*, H. Takagi†, B. Batlogg*, H. W. Zandbergen‡, J. J. Krajewski*, W. F. Peck Jr*, R. B. van Dover*, R. J. Felder*, T. Siegrist*, K. Mizuhashi†, J. O. Lee†, H. Eisaki†, S. A. Carter* & S. Uchida†

* AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

† Department of Applied Physics, University of Tokyo, Tokyo 113, Japan

‡ National Centre for High Resolution Electron Microscopy, Technical University, Delft, The Netherlands

COPPER oxide compounds have dominated superconductivity research since 1986 because of their very high transition temperatures (T_c s). In contrast, no new families of high- T_c intermetallic compounds have been discovered since the A15-type Nb_3X compounds were first reported in 1953¹. The intermetallics with highest T_c s have all been based on niobium, with the highest T_c s being 20.7 K for bulk Nb_3Ga and 23.2 K for sputtered films of Nb_3Ge (refs 2, 3). Here we report the observation of superconductivity at 23 K in a multiple-phase bulk sample of a quaternary intermetallic, yttrium palladium boride carbide. This is higher than any T_c reported previously for a bulk intermetallic compound. Although the materials are not yet single-phase, the superconducting volume fraction is large. We propose that this compound may represent the first of a new family of superconducting intermetallics with relatively high T_c s.

It has been widely speculated that the low mass of boron should be conducive to the occurrence of high phonon frequencies and consequently high T_c s, and indeed many boride superconductors are known. Generally, however, these have low T_c s, in the 2–7 K range⁴. The highest T_c s previously observed for intermetallic borides are 11.7 K for $LuRh_4B_4$ (refs 5, 6), and recently 12 K at a low volume fraction in samples of yttrium nickel boride⁷.

Our samples were prepared by arc-melting. Starting materials were Y shavings (99.9% purity), Pd powder (>99.99% purity), and coarse powders of C (99.99% purity) and B (99.6% purity). Samples of 0.6–0.75 g total weight were first pressed into 0.25-inch pellets. They were then arc-melted under Ar on a standard water-cooled copper hearth two or three times, with the button turned over between each melt. Our investigation of the properties of arc-melted samples in the Y–Pd–B–C quaternary system showed that all four elements were required to be present for

the observation of bulk superconductivity at 23 K. Superconductivity was observed for a narrow range of compositions at relatively low carbon contents. Very large magnetic shielding fractions, near 50%, and large magnetic flux exclusion fractions, 10–15% of that expected for perfect diamagnetism, were obtained from the samples for compositions of the form $YPd_5B_3C_x$, for $0.3 \leq x \leq 0.4$. Annealing the superconducting samples for 2 d at 900 °C (sealed evacuated quartz tubes, Ta-foil wrapped) resulted in the complete loss of superconductivity.

The temperature dependence of the magnetization measured in a field of 20 Oe in a SQUID magnetometer (Quantum Design) for an arc-melted sample of composition $YPd_5B_3C_{0.3}$ is given in Fig. 1. The sample cooled in zero field showed a magnetic shielding approximately equal to 40% of that expected for perfect diamagnetism. On cooling in the field, the flux expulsion (Meissner effect) corresponds to ~10% of that expected for perfect diamagnetism, showing that the superconductivity is indeed a bulk effect. The existence of hysteresis between zero-field cooling and field cooling indicates that the new compound is a type II superconductor. The inset to Fig. 1 shows, on an expanded scale, the region around T_c , revealing a critical temperature of 22.6 K. Similar T_c s were obtained for other samples in the optimal composition range.

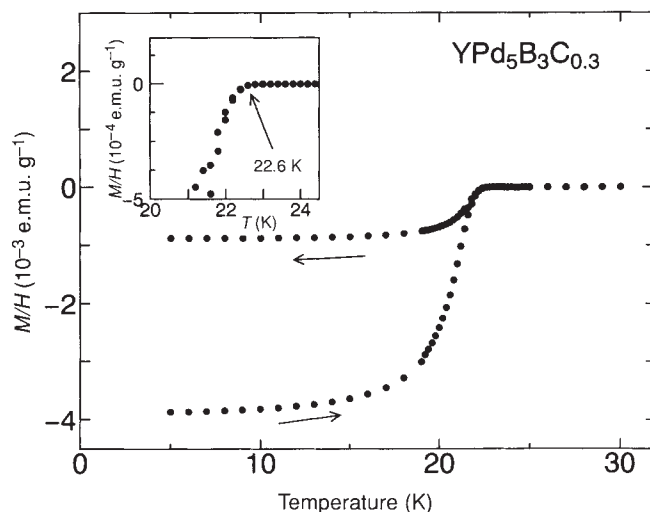


FIG. 1 Temperature-dependent magnetization M , in a field H of 20 Oe, for $YPd_5B_3C_{0.3}$. The onset of superconductivity, shown in the inset, is 22.6 K. The data were obtained both for warming after zero-field cooling, and for cooling in the field, as denoted by arrows.

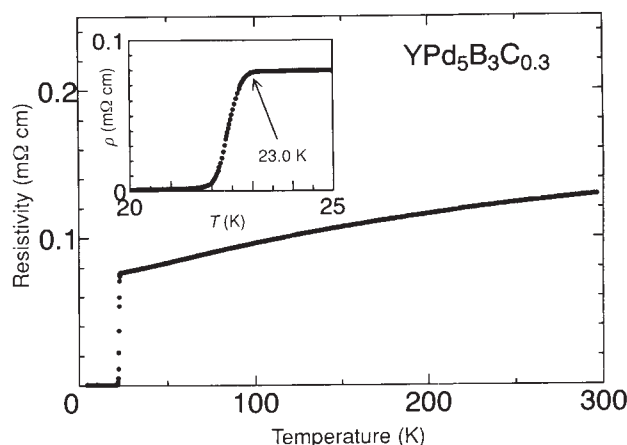


FIG. 2 Temperature-dependent resistivity ρ for sample of Fig. 1. Inset, expanded view showing the superconducting transition.

Figure 2 shows the temperature dependence of the resistivity for a (polycrystalline) sample of composition $\text{YPd}_5\text{B}_3\text{C}_{0.3}$. The normal-state resistivity is of the order of $100 \mu\Omega \text{ cm}$ at room temperature, and shows a slight 'S'-shaped temperature dependence. Of course characterization of the actual normal-state properties of the superconductor will require single-phase polycrystalline materials or single crystals. The inset to Fig. 2 shows the resistive superconducting transition in more detail. The 10–90% transition width is $\sim 0.5 \text{ K}$, and the onset temperature is 23.0 K .

An alternative method for synthesis of the yttrium palladium boride carbide was to start with the ternary (omitting C) or quaternary alloys and arc-melt them on a carbon block (volume $\sim 2 \text{ ml}$) laid on the copper hearth. This allowed experimentation with slow cooling of the samples as the heated carbon block cooled relatively slowly, but it did not allow control of the carbon content of the resulting samples, as they reacted with the block. A sample made in this manner by melting a pre-melted ternary alloy of composition $\text{Y}_{1.25}\text{Pd}_{4.5}\text{B}_{4.25}$ had a slightly higher magnetically measured superconducting transition temperature of 23.0 K . The temperature-dependent magnetization data for this sample are shown in Fig. 3. The shielding and flux exclusion signals correspond to $\sim 20\%$ and 5% of perfect diamagnetism, respectively. The inset to Fig. 3 shows in detail the region near T_c . We have not yet been able to obtain this T_c in a sample of controlled carbon content by conventional arc melting. Factors such as carbon content and the amount of time spent at particular temperatures during the slower cool are likely to affect T_c , and need further study.

Superconductivity in small quantities can also be observed in 'ternary' Y–Pd–B alloys with no intentionally added carbon. The temperature dependence of the magnetization at 20 Oe for one of the best 'ternary' samples (synthesized by conventional arc melting) at composition $\text{Y}_{1.5}\text{Pd}_{4.5}\text{B}_4$, is shown in Fig. 4. The magnitude of the shielding and flux exclusion signals correspond roughly to 0.8% and 0.4% of perfect diamagnetism, respectively, suggesting that the superconducting phase is a minor phase with a volume fraction of the order of 1% . This sample shows only a 10% resistance drop at T_c , shown in the inset to Fig. 4. The superconductivity may be present in the ternary alloys because the superconducting phase as a pure boride is highly unstable, but it is much more likely that the carbon is inadvertently present in small amounts in the 'ternary' alloys, perhaps from a carbon impurity in the boron starting material.

Superconducting compositions with $T_c > 22.5 \text{ K}$ in the Y–Pd–B–C system are, initially at least, harder to obtain, reminiscent

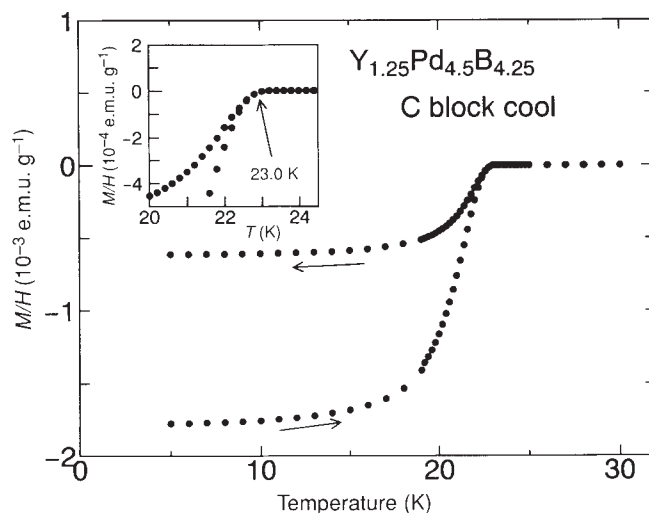


FIG. 3 Temperature-dependent magnetization M , in a field H of 20 Oe , for $\text{Y}_{1.25}\text{Pd}_{4.5}\text{B}_{4.25}\text{C}_x$ prepared by melting on a heated carbon block. The onset of superconductivity, shown in the inset, is 23.0 K . The data were obtained both for warming after zero-field cooling, and for cooling in the field, as denoted by arrows.

of the highest- T_c compositions for A15 superconductors⁸. The equilibrium Y–Pd–B and Y–Pd–B–C phase diagrams have not been studied, and there are apparently no known ternary Y–Pd–B or quaternary Y–Pd–B–C compounds⁹. It remains to be seen whether the new superconductor is a true quaternary compound, containing structurally and chemically distinct B and C atoms, or whether there is a B–C solid solution in a Y–Pd–B ternary, with the ratio B/C changing the electron count, much as Sr does in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. Exact determination of the superconductor composition may await single-crystal structure determination, as intermetallic phases can be complex. We believe that the yttrium palladium boride carbide superconductor will prove to be the first of a new family of high- T_c intermetallic compounds, and we

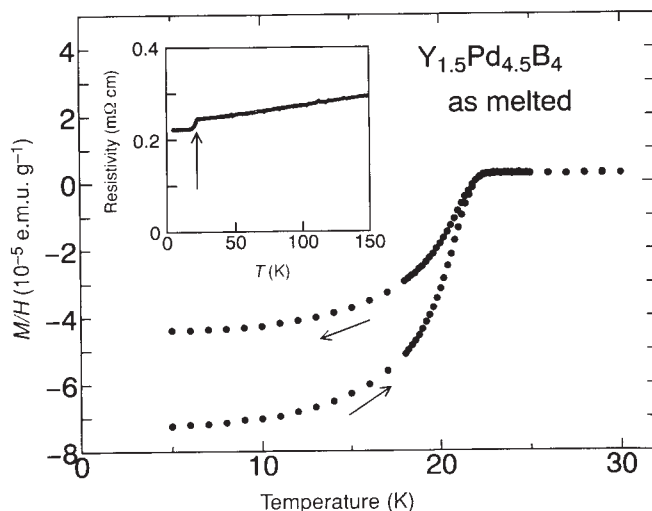


FIG. 4 Temperature-dependent magnetization M , in a field H of 20 Oe , for $\text{Y}_{1.5}\text{Pd}_{4.5}\text{B}_{4.25}$ prepared by conventional arc melting. Inset shows the resistivity for this sample, with a small resistivity drop at T_c . The data were obtained both for warming after zero-field cooling, and for cooling in the field, as denoted by arrows.

suggest that boride-carbides (and borides) represent one road to high T_c that is worthy of further exploration. □

Received 13 December; accepted 21 December 1993.

- Hardy, G. F. & Hulm, J. D. *Phys. Rev.* **89**, 884–885 (1953).
- Gavaler, J. R., Janocko, M. A. & Jones, C. J. *Appl. Phys.* **45**, 3009–3013 (1974).
- Testardi, L. R., Wernick, J. H. & Rogers, W. H. *Solid St. Commun.* **15**, 1–4 (1974).
- Roberts, B. W. J. *Phys. Chem. Ref. Data* **5**, 581–794 (1976).
- Johnston, D. C. *Solid St. Commun.* **24**, 699–702 (1977).
- Matthais, B. T., Corenzwit, E., Vandenberg, J. M. & Barz, H. E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1334–1335 (1977).
- Mazumdar, C. et al. *Solid. St. Commun.* **87**, 413–416 (1993).
- Flükiger, R. in *Concise Encyclopedia of Magnetic and Superconducting Materials* (ed. Evetts, J.) 1–15 (Pergamon, Oxford, 1992).
- Villars, P. & Calvert, L. D. (eds) *Pearsons Handbook of Crystallographic Data for Intermetallic Phases* 2nd edn (ASM Int., Materials Park, Ohio, 1991).

ACKNOWLEDGEMENTS. The work at the University of Tokyo was supported by the Ministry of Education, Science and Culture of Japan.

Role of sp^3 defect structures in graphite and carbon nanotubes

H. Hiura, T. W. Ebbesen*, J. Fujita, K. Tanigaki & T. Takada

Fundamental Research Laboratories, NEC Corporation, 34 Miyukigaoka, Tsukuba 305, Japan

THE discovery of fullerenes^{1,2} and carbon nanotubes^{3,4} has led to the realization that it should be possible to tailor the properties of graphitic sheets if their geometry can be controlled^{5–7}. In exploring these possibilities, we have found that the folding and tearing of graphitic sheets follow well defined patterns which seem to be governed by the formation of sp^3 -like line defects in the sp^2 graphitic network. Our studies with the atomic force microscope and scanning tunnelling microscope reveal that these folds and tears occur preferentially along the symmetry axes of graphite, and that 'ripples' are observed in the curved portions of the folds. We also see ripples in deformed carbon nanotubes. They lie along the directions for which sp^3 -like line defects can form most easily to relieve strain. Our *ab initio* molecular orbital calculations indicate that the ripples stabilize the π -electronic energy in the bent structures with the total energy balance being determined by the amount of nuclear repulsion. These results provide insight into the geometries that graphitic structures will preferentially accommodate, and the properties that might ensue.

Highly oriented pyrolytic graphite (HOPG) was freshly cleaved and imaged by atomic force microscopy (AFM) and scanning tunnelling microscopy (STM) as shown in Fig. 1. Figure 1a (low magnification AFM) shows two steps running parallel to each other. Each step has a 3.4 Å height indicating that they correspond to a single graphitic sheet. Along the step edge, we see polygonal structures formed by tearing and folding of the graphitic sheets. The graphitic sheets are cut, folded and refolded at very specific angles to the step edge: 30°, 60°, 90° and 120°, corresponding to the angles between the symmetry axes of graphite. The sixfold symmetry is expected due to the honeycomb structure of the graphitic sheet which has three axes (I ([100]), I' and II'') at 60° to each other and three other equivalent axes (II ([210]), II' and II'') separated from the first set by 30°. Figure 1b shows very clearly an example of massive tearing along the angles defined above. In nearly all of the cases that we have observed so far, tearing and folding occurs at angles corresponding to the difference between two or more symmetry axes. These directions must be preferred for energetic reasons.

Some of the structures in Fig. 1a are shown at higher mag-

nification in Fig. 1c, d and e. The most striking aspect of Fig. 1c and d are the ripples along the folds. These ripples are all parallel to each other and to the direction of the fold, and they are most apparent in the most curved part of the fold. At the normal step edge, where the graphite sheet is cut, no ripples are observed (Fig. 1e). The AFM scanning was carried out under different conditions and at different angles to the ripples to confirm that these structures are intrinsic and real. Similar images are also obtained by STM (Fig. 1f). These control experiments make it most unlikely that the ripples are due to tip structure or other artefacts.

Cross-sectional analysis of the fold shown in Fig. 1c reveals that it consists of a single sheet of graphite (3.4 Å), folded over with ripples spaced by 10–15 Å. The triangular structure shown in Fig. 1d is made from the fold in Fig. 1c that has been refolded. The observed height of the ripples is ~1 Å, which is probably exaggerated by the response of the AFM. We believe that the true height is probably smaller, of the order of a few tenths of an ångström. The extremely small size of the ripples indicates that they have their origin directly in the atomic structure of the graphite sheet.

In general, the smaller the radius of the curvature, the more pronounced are the ripples. Sometimes, due to the force of the AFM tip, one can observe that a smooth fold of a graphitic sheet becomes more and more flattened by the scanning. As the flattening increases, ripples appear and become increasingly pronounced where the strain is the greatest in the fold.

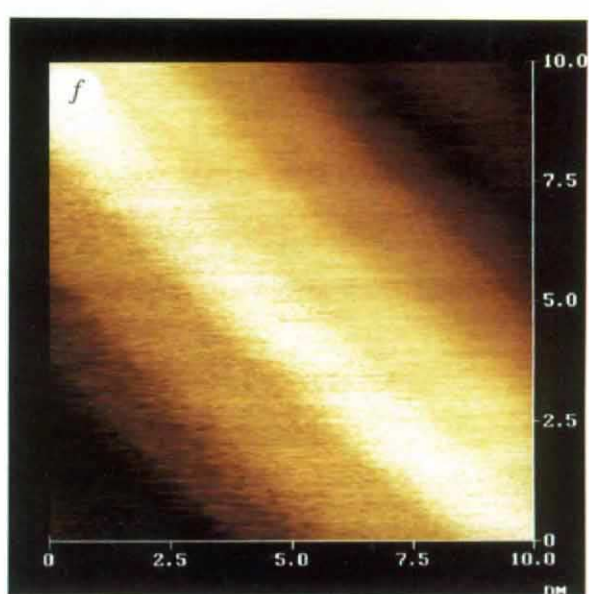
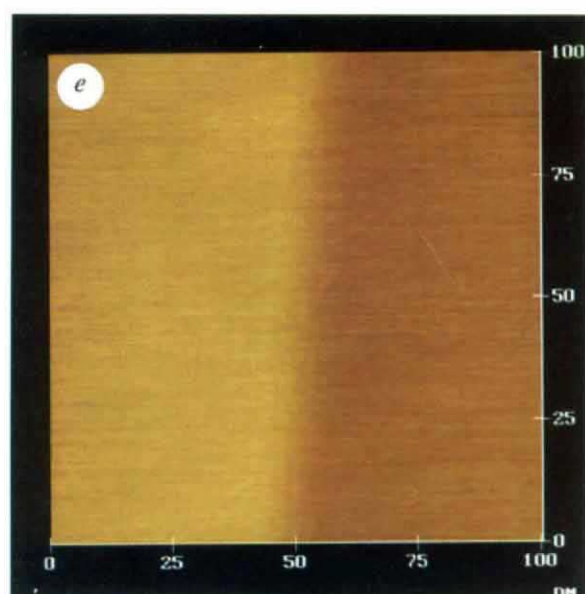
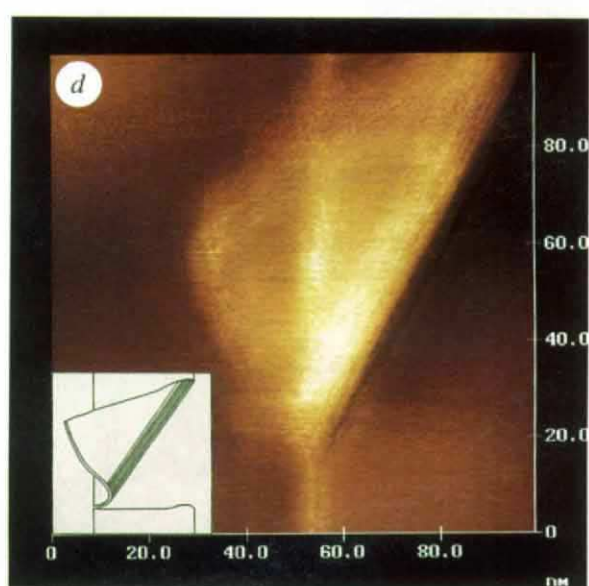
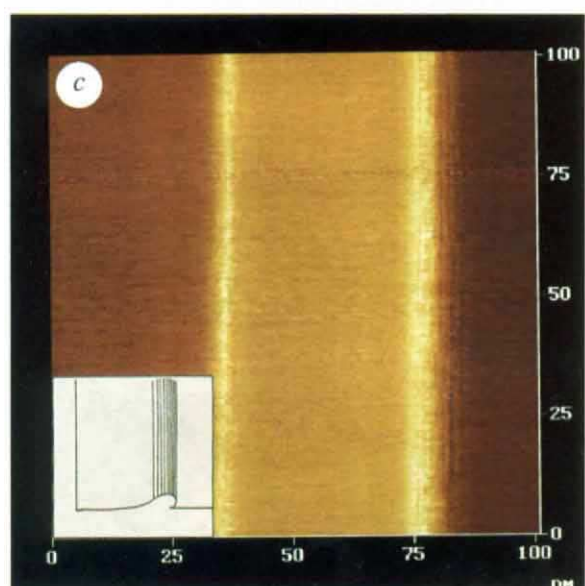
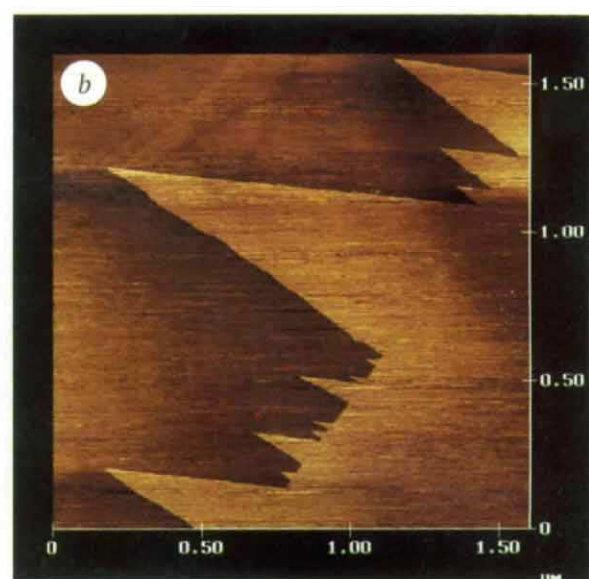
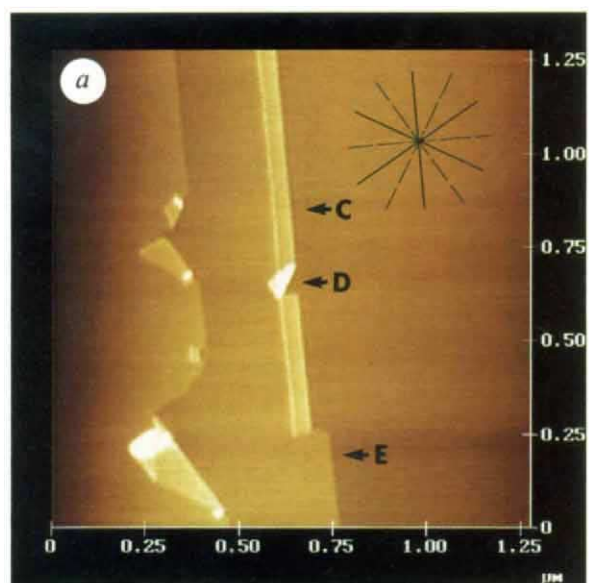
Nanotubes, prepared by the carbon arc method^{4,8}, also seem to be occasionally deformed (flattened) by the force of the AFM tip, resulting in rippled surfaces as shown in Fig. 2a. These ripples are overall parallel to the nanotube direction, although closer inspection (Fig. 2b) reveals the interplay between the ripples and carbon atom arrays. In this case the carbon atoms are arrayed at ~15° to the long axis of the nanotube, reflecting its helicity³. The ripple spacing is relatively random ranging from 5 to 20 Å (the dark and light zones running vertically in Fig. 2b). These values correspond to 2–8 times the periodicity (2.46 Å) of the flat hexagonal carbon structure. Flattening was observed only for tubes with diameter <~10 nm. We believe that the smaller nanotubes, containing a smaller number of concentric layers, are weaker and therefore are more easily deformed. Obviously all the layers of nanotube must become deformed to some extent when a tube is flattened. That large cylindrical carbon whiskers can be deformed by mechanical pressure was reported by Bacon in 1960 (ref. 9). More recently, it has been found that nanotubes can be deformed simply by van der Waals forces¹⁰.

These observations suggest that the ripples serve to relieve strain both in folded graphitic sheets and deformed nanotubes, and thereby to stabilize the new structures. These bent structures cannot be explained in terms of the ordinary planar sp^2 hybrid carbon-carbon bonding of the flat graphitic sheet. When the graphitic sheet bends, the bonding must lose some of its sp^2 character and gain some sp^3 character. So the ripples can be seen as sp^3 -like line defects in the sp^2 graphitic sheets.

The change from sp^2 - to sp^3 -like character must always involve a pair of carbon atoms, as it is the π -bonding that is being disrupted. This simple notion has immediate consequences for

FIG. 1 a, Surface of highly oriented pyrolytic graphite (HOPG) from atomic force microscopy (AFM). The symmetry axes fall into two sets of equivalent directions (—) and (---), one set belonging to the [100] direction and the other to the [210] direction (atomic imaging did not enable us to specifically assign each direction). Arrowheads labelled C, D and E indicate the features shown at higher magnification in c, d and e. b, Massive tearing of single graphite sheet from scanning tunnelling microscopy (STM). c, d and e, Details of a showing fold c, triangular fold (d) and graphite step (e). Notice the ripples in the folds in images c and d but not in the step e. f, An STM image of ripples in a fold. (Note that the units on the axes are μm in a and b, and nm in c–f.)

* To whom correspondence should be addressed.



the type of sp^3 -like line defects that can form, and therefore for the directions of tearing and folding that should be observed. Figure 3 shows the computer-generated images of the possible line defects, which can be classified into two groups according to the symmetry axis they follow. The first group follows the I ([100]) set of equivalent symmetry directions, and the line defects may involve two different possible carbon atom pairs. There are two conformations for each pair, either chair (Ic and Ic') or boat (Ib and Ib') as shown in Fig. 3. For the second group, which follows the II ([210]) set of equivalent symmetry

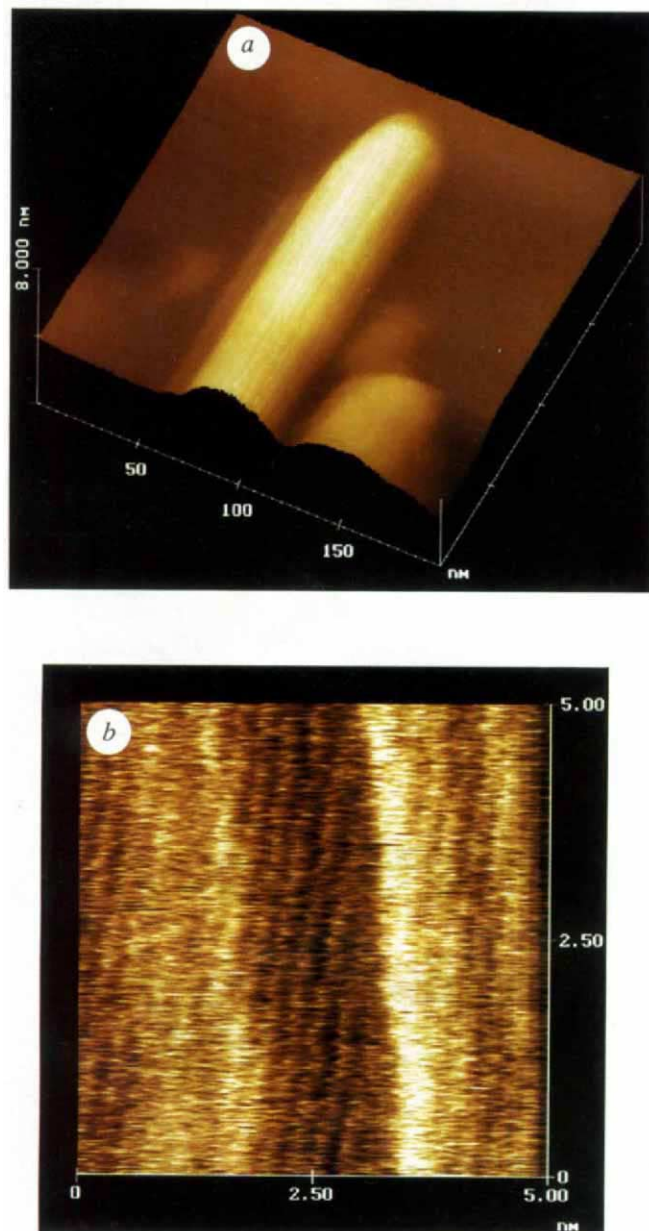


FIG. 2 *a*, Ripples on the surface of a nanotube prepared by the carbon arc method^{4,8}. *b*, Details of the surface showing the relationship between the ripples parallel to the long axis of the nanotube (vertical direction) and the carbon atom array at $\sim 15^\circ$ tilt. We have observed by AFM >100 nanotubes with diameters ranging from 10 to 500 Å, and found that when the diameter is relatively large ($> \sim 100$ Å) the nanotubes have no ripples on the surface⁸. Further, ripples are only observed on the smaller nanotubes when the AFM measurements (of the tube width and height) indicate that the nanotube is not circular but has an oval cross section. (Note that the units on the axes of *a* and *b* are nm.)

directions, there is one conformation for one pair of carbon atoms (IIa in Fig. 3) and two conformations for the other pair (boat IIb and chair IIc). In summary, sp^3 -type line defects can occur along the symmetry axes involving four different sets of carbon pairs in the sp^2 hexagonal network, resulting in seven possible conformations. These basic line defects have resonance structures, but their discussion would be beyond the scope of this Letter.

From the above it is clear why folding and rippling occur preferentially along the symmetry axes and result in the structures seen in Fig. 1. Along these axes, the sp^3 -like line defects form most readily and therefore are energetically preferred over some intermediate angles. For the same reasons, tearing should also be favoured along these axes because this process involves complete sp^2 bond breaking.

To investigate how formation of a ripple relieves the strain in a graphitic sheet, we have performed *ab initio* calculations (MINI-4¹⁴) on a piece of graphitic sheet composed of 16 hexagons (60 carbon atoms). This sheet was curved as if it was a part of the surface a tube with 2-nm radius, and its total energy was calculated. Then different types of sp^3 -like line defects, according to our classification above, were introduced into the sheet in order to replace the smooth curvature by bends and flat areas. The *ab initio* calculations on these structures indicate that the π -electronic energy is stabilized by the introduction of bends

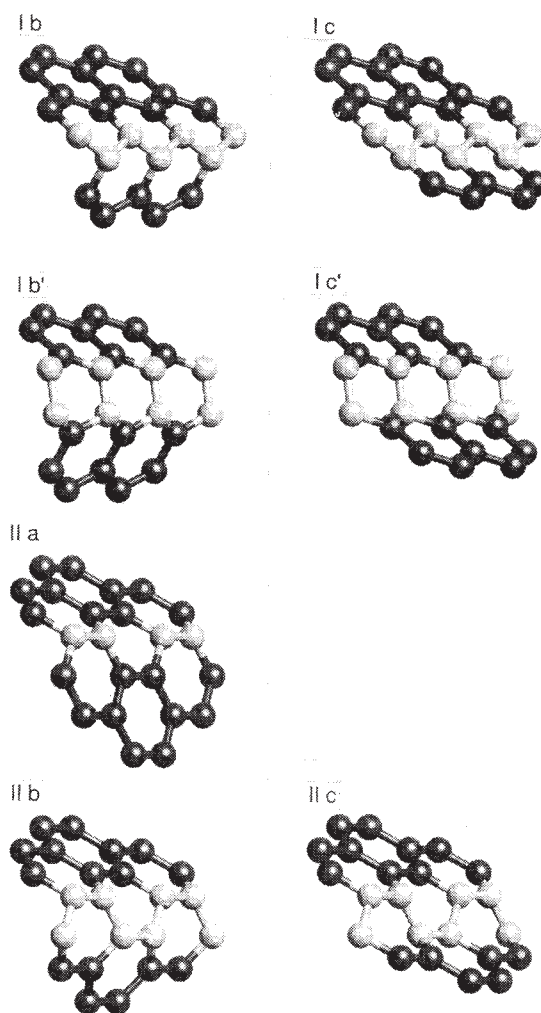


FIG. 3 Computer-generated illustrations of the basic sp^3 -like line defects (grey) that can occur in an sp^2 network (black) of carbon atoms. I and II indicate the symmetry axes, [100] and [210] respectively, to which these line defects belong.

into the originally strained sheet. But the total energy balance is determined by whether the increase in the coulombic (nuclear) repulsion is smaller or larger than the gain in π -energy. In the non-optimized geometries we tried, the total energy balance was unfavourable because of the coulombic repulsion (from ~ 10 to 0.6 kcal per mole of carbon atoms). In the best case, we calculated the energy loss to be only 3.6 kcal mole $^{-1}$ per benzenic ring due to the introduction of bends. This is very small considering that the structure (bond lengths) was not optimized in these calculations and supports the general notion that sp^3 -like line defects should in principle be able to reduce strain in a curved graphitic sheet. Although it is a very different situation, it is interesting to note that in aromatic carbon polymers, single-double bond alternation is believed to be more stable than an isotropic distribution of the π -density in the hexagonal polymer network¹¹.

The sp^3 -like line defects should result in well defined edges in strongly curved graphitic sheets. Such edges can actually be seen in transmission electron micrographs of ribbon edges of filamentous graphite¹² and other folded sheets of graphite¹³. These are not due to the presence of pentagons as the structures are apparently not closed. In contrast, closed structures such as nanoparticles and the tips of nanotubes have well defined edges due to the presence of pentagons. At the same time, it is possible that line defects may stabilize these structures.

The introduction of sp^3 -like defects must significantly change

the properties of the surrounding sp^2 network. For instance, electronic properties should be affected although the amount of change will probably depend on the geometry of both the defect and the overall structure. One can easily imagine idealized wires made of carbon sheet with conductive zones separated by ripples or bends. The sp^3 -like line defects would act as non-conducting barriers separating structures in the overall carbon network. Further theoretical and experimental studies should tell what to expect and what is possible in such cases. $\square$

Received 2 August; accepted 9 November 1993.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–164 (1985).
2. Kratschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
3. Iijima, S. *Nature* **354**, 54–58 (1991).
4. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
5. Mackay, A. L. & Terrones, H. *Nature* **352**, 762 (1991).
6. Mintmire, J. W., Dunlap, B. I. & White, C. T. *Phys. Rev. Lett.* **68**, 631–634 (1992).
7. Hamada, N., Sawada, S. & Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
8. Ebbesen, T. W. et al. *Chem. Phys. Lett.* **209**, 83–90 (1993).
9. Bacon, R. J. *Appl. Phys.* **31**, 283–290 (1960).
10. Ruoff, R. D., Tersoff, J., Lorents, D. C., Subramoney, S. & Chan, B. *Nature* **364**, 514–516 (1993).
11. Tanaka, K., Ohzeki, K., Nankai, S. & Yamabe, T. *J. Phys. Chem. Solids* **44**, 1069–1075 (1983).
12. Murayama, H. & Maeda, T. *Nature* **345**, 791–793 (1990).
13. Ugarte, D. *Chem. Phys. Lett.* **198**, 596–602 (1992).
14. Huzinaga, S. *Gaussian Basis Sets for Molecular Calculations* (Elsevier, Amsterdam, 1984).

ACKNOWLEDGEMENTS. We thank S. Handa for his assistance with the computer graphics.

Formation of an ordered Langmuir monolayer by a non-polar chain molecule

Mengyang Li, Anibal A. Acero, Zhengqing Huang & Stuart A. Rice

Department of Chemistry and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA

AMPHIPHILIC molecules, which contain a polar 'head' group and a hydrophobic 'tail', will generally form monolayers at the air–water interface in which the molecules protrude from the interface with the tails in the air. These 'Langmuir monolayers' exhibit a variety of phases as a function of surface pressure and temperature, including an ordered, two-dimensional crystalline phase^{1–3}. It might be considered that the amphiphilic nature of the molecules allows this oriented, tail up arrangement and the consequent formation of a well ordered phase. Here, by contrast, we report that the non-polar molecule perfluoro-*n*-eicosane ($F(CF_2)_{20}F$) will form stable, ordered Langmuir monolayers on water, even though they have no amphiphilic character. Grazing-incidence X-ray diffraction studies of these monolayers show that the molecules are vertically aligned and are packed in a hexagonal array over a range of temperature. We suggest that van der Waals forces alone are sufficient to stabilize the monolayer in this case.

Our sample of perfluoro-*n*-eicosane (>97% pure, from PCR Inc., Gainesville, Florida) was used as received. Solutions were prepared by dissolving a very small amount of the compound, typically 50 mg, in 25 ml of a 4:1 mixture of 1,1,2-trichlorotrifluoroethane and perfluoropentane. A monolayer was prepared by spreading the solution onto the water surface using a microsyringe and allowing the solvent to evaporate. Isotherms (surface pressure against surface area) of the perfluoro-*n*-eicosane monolayer are shown in Fig. 1. A conventional linear extrapolation of the high-pressure limb of the isotherm to zero pressure yields

a limiting area of ~ 30 Å² per molecule, very close to the area per molecule in crystalline perfluoro-*n*-eicosane (28.1 Å² per molecule)^{4,5}, and therefore consistent with the existence of a monolayer. In the low surface-pressure region (< 2 dyne cm $^{-1}$), where we have carried out all our measurements, the surface pressure is time-independent and the isotherms are reversible within the precision of our measurements.

The grazing-incidence X-ray diffraction apparatus has been described elsewhere⁶. It consists of a sample chamber with a 180° gapped (gas filled) double Mylar window for X-ray access, a Teflon Langmuir trough with isoperimetric Teflon ribbon, and associated temperature control, surface density control and surface-pressure measurement equipment. The temperature of the sample substrate (water) was regulated (± 0.05 °C) by a dual circulator system which separately controlled the temperatures of the cover and the base of the apparatus. The surface pressure was determined using the Wilhelmy plate method⁷, with a precision of ± 0.05 dyne cm $^{-1}$.

Our X-ray source was the wiggler line (hutch A1) at the Cornell high energy synchrotron source (CHESS). A 10-cm Pt-on-glass mirror was used to deflect the monochromatic beam of 8.2405-keV (1.5056-Å) X-rays downwards so that it intersected the monolayer surface at an angle of 1.79 mrad, which is slightly smaller than the critical angle for total external reflection. The intensity of the incident X-ray beam was continuously monitored by a nitrogen-filled ion chamber. The diffracted radiation was collected by a position-sensitive linear detector oriented with its long axis perpendicular to the plane of the water–monolayer interface. This detector, which has 500 resolution elements, was located so as to have a maximum z -axis angular resolution of 0.0039 Å $^{-1}$ at 1.0 Å $^{-1}$. The in-plane (plane of the water–monolayer interface) angular resolution of the detector was determined by Soller slits; the plate separation of these slits provided an in-plane angular resolution of 0.01 Å $^{-1}$ at 1.0 Å $^{-1}$.

The intensity of the diffracted X-rays measured at each angle in-plane and out-of-plane was normalized with respect to the incident intensity. Because of the limited duration of a fill of the CHESS storage ring and the long time required to carry out a

scan, the diffraction pattern of the monolayer has been pieced together from the results of many scans. Unfortunately, because of the large projection of the X-ray beam on the sample surface, very small variations in the beam location from fill to fill of the storage ring generate uncertainties which prevent the intensities of the peaks measured in different scans from being compared quantitatively.

In-plane diffraction patterns were constructed from the data set by integrating, for each in-plane angle, over the range $0 < Q_z < 0.5 \text{ \AA}^{-1}$, where Q_z is the momentum transfer perpendicular to the interface. The out-of-plane scattering intensity, along the Bragg rod, was stored by compressing the 500-pixel resolution of the detector into 200 bins, thereby reducing the maximum out-of-plane resolution to $0.01 \text{ \AA}^{-1}$. For purposes of analysis, the out-of-plane data were treated as having a resolution of $\Delta Q_z = 0.1 \text{ \AA}^{-1}$, obtained by integration over ~ 10 of the stored resolution elements (note that the actual number of resolution elements summed depends on the out-of-plane angle). The background scattered X-ray intensity was assumed to vary linearly across a diffraction peak, and subtracted from the peak. Both gaussian and lorentzian line shapes were used to fit the corrected peak shapes; the gaussian fits generally had a smaller χ^2 -statistic. The scattering intensity was determined by integrating the area under the background-corrected peak that fitted the data best, using the χ^2 -statistic as the criterion of best fit.

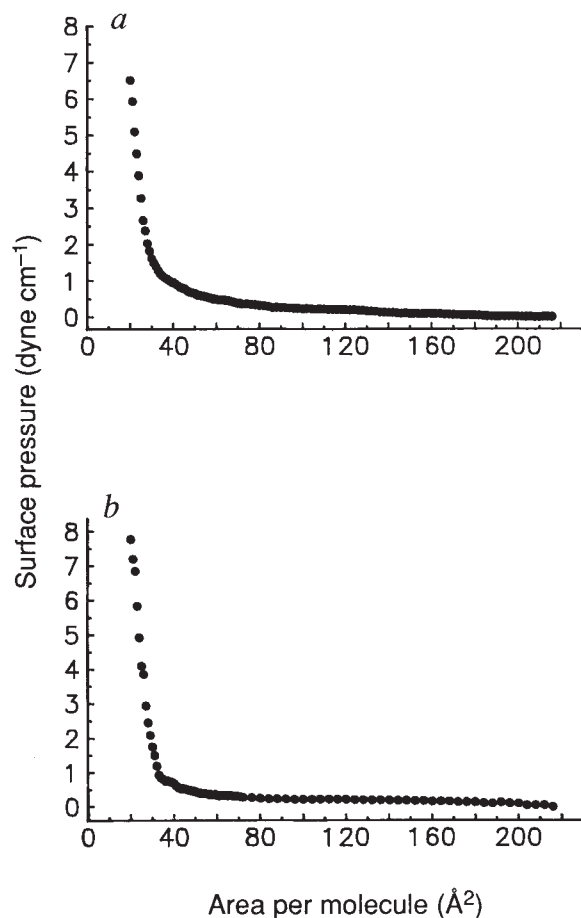


FIG. 1 Isotherms of the perfluoro-*n*-eicosane monolayer on water at 4 °C (a) and 20 °C (b). The surface pressure was measured with a precision of $\pm 0.05 \text{ dyne cm}^{-1}$ and an accuracy of $\pm 0.5 \text{ dyne cm}^{-1}$. The rate of monolayer compression varied along the isotherm to allow for surface pressure relaxation; the compression rate used was everywhere $< 2 \text{ \AA}^2 \text{ min}^{-1}$ per molecule.

TABLE 1 Structural parameters for a monolayer of $\text{F}(\text{CF}_2)_{20}\text{F}$ on water

Temp. (°C)	Area ($\text{\AA}^2$ per molecule)	Q_1^* ($\text{\AA}^{-1}$)	$\Delta Q_1^\dagger$ ($10^{-3} \text{ \AA}^{-1}$)	Q_2^* ($\text{\AA}^{-1}$)	$\Delta Q_2^\dagger$ ($10^{-3} \text{ \AA}^{-1}$)	Q_3^* ($\text{\AA}^{-1}$)	$\Delta Q_3^\dagger$ ($10^{-3} \text{ \AA}^{-1}$)
4	1082	1.2829	9.6 $\ddagger$	—	—	—	—
4	250	1.2828	10.2	—	—	—	—
4	100	1.2797	8.8	—	—	—	—
4	60	1.2823	8.7	2.2114	10.0	2.5531	10.1
4	60	1.2822	8.1 $\ddagger$	2.2134	15.1	—	—
4	40	1.2825	8.2	2.2109	15.3	—	—
4	33	1.2820	8.2	—	—	—	—
4	30	1.2818	8.2	—	—	—	—
4	28.6	1.2820	8.3	—	—	—	—
20	32	1.2808	9.1	—	—	—	—
20	32	1.2805	8.7	—	—	—	—
25	1400	1.2781	8.0 $\ddagger$	—	—	—	—
25	500	1.2779	8.0	—	—	—	—
25	400	1.2788	8.1	—	—	—	—
25	400	1.2761	8.3	—	—	—	—
25	400	1.2785	8.6	2.2057	11.0	—	—
25	300	1.2782	7.9	—	—	—	—
25	42.7	1.2777	8.2	2.2045	8.2	—	—
25	29.7	—	—	2.2050	9.6	—	—

These structural parameters are based on gaussian fits to the grazing incidence X-ray diffraction data.

* In-plane diffraction peak position.

$\dagger$ Diffraction peak width (half-width at half-maximum).

$\ddagger$ Width inaccurate due to large step used in scan or distorted peak shape.

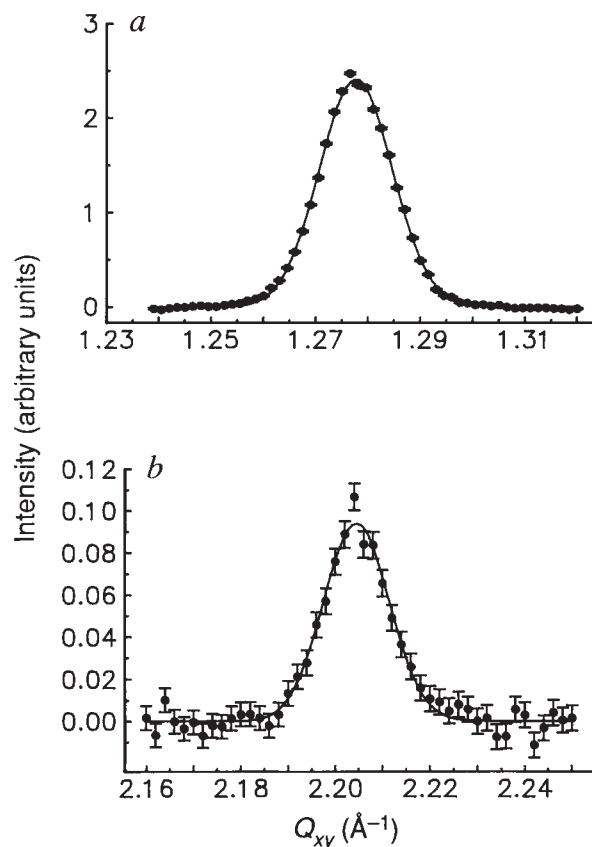


FIG. 2 In-plane diffraction peaks (near $Q_z = 0$) of the perfluoro-*n*-eicosane monolayer at 25 °C on water at a reciprocal surface density of $42.7 \text{ \AA}^2$ per molecule. The dots with statistical error-bars represent the experimental data. The lines are gaussian fits to the experimental data. a, First-order diffraction peak. b, Second-order diffraction peak. ($Q_{xy} = (Q_x^2 + Q_y^2)^{0.5}$ is the magnitude of the in-plane momentum transfer).

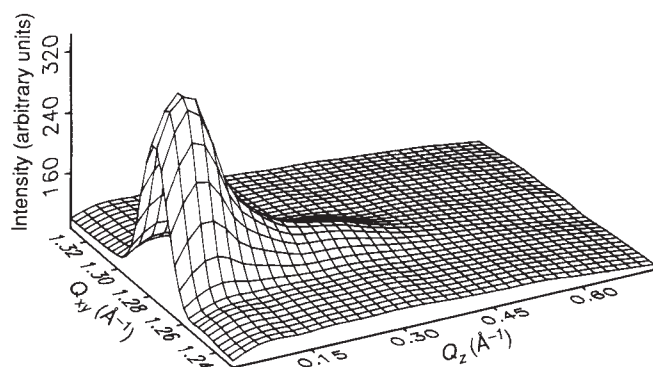


FIG. 3 A three-dimensional plot of the diffraction data in Q_x , Q_z reciprocal space for the same monolayer as described in the caption to Fig. 2. The first-order diffraction peak is shifted away from $Q_z=0$ by $0.06 \text{ \AA}^{-1}$, which is the accuracy with which we were able to determine the absolute zero of the out-of-plane momentum transfer. The data show that the collective tilt of the perfluoro-*n*-eicosane molecules is $<3^\circ$ (measured from the normal to the surface).

Typical diffraction data are shown in Figs 2 and 3; a summary of our results is displayed in Table 1. At both 4 and 25 °C the positions of the diffraction peaks are independent of surface coverage over the range ~ 29 to $\sim 1,000 \text{ \AA}^2$ per molecule, which we interpret as arising from the coexistence of a structured condensed phase and an undetected (gas) phase, hence as evidence for a first-order transition between those phases: the correlation range is of the order of $750 \text{ \AA}$. We note that the three diffraction peaks found have reciprocal lattice locations which have the relationship Q_1 , $\sqrt{3}Q_1$ and $2Q_1$ (Table 1) ($Q=4\pi/\lambda \sin(\theta/2)$, where θ is the scattering angle), and that the structure function in the Q_z direction is peaked very near $Q_z=0$ (Fig. 3). Our diffraction data are consistent with a monolayer structure with hexagonal packing of vertical perfluoro-*n*-eicosane molecules on the water surface; the unit cell parameters are (at 4 °C) $a=b=5.658 \text{ \AA}$, $\gamma=120^\circ$ and at 25 °C are $a=b=5.677 \text{ \AA}$, $\gamma=120^\circ$. We also note that the full width at half maximum (FWHM) of the rod scan, taken in the cut through the peak shown in Fig. 3, is $0.21 \text{ \AA}^{-1}$. The perfluoro-*n*-eicosane molecule length, estimated from $L=2\pi/\Delta Q_z$ (FWHM)⁴, is $29.9 \text{ \AA}$, to be compared to the value $28.3 \text{ \AA}$ reported in ref. 1. The inferred molecular length is consistent with the existence of a monolayer with close-packed vertical molecules.

Our data show that at all surface coverages the nearest-neighbour spacing is consistently slightly larger at 25 °C than at 4 °C (5.677 and $5.658 \text{ \AA}$ respectively). We have fewer data points at 20 °C, but those available yield a nearest-neighbour spacing of $5.664 \text{ \AA}$. The change in nearest-neighbour spacing on raising the temperature of the monolayer from 4 to 20 °C is about that expected, whereas the change from 20 to 25 °C is rather large to be attributed entirely to thermal expansion. We note that in Teflon there is a phase transition at 30 °C in which the molecules are displaced longitudinally by irregular amounts⁸, and in which there is a small increase in the lattice spacing without change in the symmetry of the projected unit cell base. Perhaps there is a similarly irregular longitudinal displacement of the molecules in the perfluoro-*n*-eicosane monolayer at 25 °C.

The observation that it is possible to form a stable Langmuir monolayer of perfluoro-*n*-eicosane on water is unusual because the molecules known to form such monolayers are almost always amphiphilic⁷; the most common of these have a polar head group and a long aliphatic-chain tail. (It is possible to generate stable monolayers of molecules which have bulk liquid crystalline phases, and which are not traditional amphiphiles with a simple alkyl chain and polar head group, for example, ref. 9. But these molecules have one or more strongly polar regions, and so are not comparable with a nonpolar adsorbate such as $\text{F}(\text{CF}_2)_{20}\text{F}$.) It is commonly believed that the polar head group in the amphiphile is necessary to anchor the molecule to the water surface, whereas the attractive interaction between the long-chain tails of the amphiphile molecules generates their self assembly⁷. But because long-chain molecules typically form

lamellar crystals with very weak interactions between the layers¹⁰, it seems plausible that, in favourable cases, the interactions within one layer are sufficiently strong to maintain the layer packing when one face is in contact with water, and that the van der Waals interactions between such a layer and the water surface are sufficient to stabilize the system as a Langmuir monolayer. □

Received 27 August; accepted 16 November 1993.

1. Möhwald, H. A. *Rev. phys. Chem.* **41**, 441–476 (1990).
2. McConnell, H. M. A. *Rev. phys. Chem.* **42**, 171–195 (1991).
3. Knobler, C. M. & Desai, R. C. A. *Rev. phys. Chem.* **43**, 207–236 (1992).
4. Schwickert, H., Strobl, G. & Kimmig, M. *J. chem. Phys.* **95**, 2800–2806 (1991).
5. Albrecht, T. et al. *J. chem. Phys.* **95**, 2807–2816 (1991).
6. Acero, A. A. et al. *J. chem. Phys.* (in the press).
7. Gaines, G. L. *Insoluble Monolayers at Gas-Liquid Interfaces* (Wiley, New York, 1966).
8. Bunn, C. W. & Howells, E. R. *Nature* **174**, 549–551 (1954).
9. Xue, J., Jung, C. S. & Kim, M. W. *Phys. Rev. Lett.* **69**, 474–477 (1992).
10. Small, D. M. *The Physical Chemistry of Lipids: Handbook of Lipid Research*, Vol. 4 (Plenum Press, New York, 1986).

ACKNOWLEDGEMENTS. We thank CHESS for time on their facility. This work was supported by the US National Science Foundation.

A high-resolution record of atmospheric CO₂ content from carbon isotopes in peat

J. W. C. White*, P. Ciais*†‡, R. A. Figge*,
R. Kenny*‡ & V. Markgraf*

* Institute of Arctic and Alpine Research and Department of Geological Sciences, Campus Box 450, University of Colorado, Boulder, Colorado 80309, USA

† NOAA/CMDL R/E/CG1, 325 Broadway, Boulder, Colorado 80303, USA

OUR understanding of how future changes in atmospheric carbon-dioxide concentrations will affect the global climate system arises in part from comparing past changes in climate and CO₂. To date, these comparisons have come mainly from ice-core data, which show a strong correlation between past atmospheric CO₂ concentration and polar temperature¹. Here we present a new method for reconstructing atmospheric CO₂ concentration using the ¹³C/¹²C ratio ($\delta^{13}\text{C}$) in mosses and sedges in peat. Our method exploits the fact that, unlike sedges and most other plants, mosses do not possess stomata, and are therefore unable to regulate their uptake of CO₂ and water. The $\delta^{13}\text{C}$ of mosses thus depends on both

‡ Present addresses: LMCE, CEA Saclay, 91191 Gif-sur-Yvette, France (P.C.); Department of Geology, New Mexico Highlands University, Las Vegas, New Mexico 87701, USA (R.K.)

atmospheric CO₂ concentration and available water, and the $\delta^{13}\text{C}$ of sedges from the same peat can be used to remove the water signal. The method provides a resolution of about a decade—much higher than is possible from ice cores. We present initial results for the past 14,000 years, which show three sharp increases in atmospheric CO₂ concentration: at 12,800 years ago, corresponding to an episode of warming in the North Atlantic region; 10,000 years ago, corresponding to the end of the Younger Dryas cold period; and 4,400 years ago, after which time modern climates were established globally.

A 10.2-m sediment core, previously studied for palaeoclimate using pollen analysis^{2,3} was collected from a raised *Sphagnum* bog near Estancia Harberton, on the Beagle Channel (54° 53' S, 67° 10' W) in southern South America. This is a cold, coastal site with an average temperature at present of ~6 °C. The low temperature ensures exceptional preservation of plant fragments. Individual plant species are separable to the base of the bog, dated at 14,000 yr BP. Sedges are present throughout the core and mosses are sparse only between 9,200 and 9,800 yr BP. Between the surface and 7,000 yr BP the dominant moss is *Sphagnum magellanicum* and below that level species of *Drepanocladus* s.l. dominate. We extracted cellulose from single-species samples representing ~10 yr (1 cm) according to standard techniques^{4,5,6}. Figure 1 shows the measurements of $\delta^{13}\text{C}_{\text{sedge}}$ and $\delta^{13}\text{C}_{\text{moss}}$. There is some common variability but also a striking contrast between the records, both in pattern and magnitude of the changes. In particular, $\delta^{13}\text{C}_{\text{moss}}$ shows very large variability, as much as 10‰, during the last deglaciation.

Sedges and mosses differ fundamentally in how they regulate carbon uptake and water loss. Sedges regulate their intrinsic water use efficiency by the opening and closing of the stomatal

pores on their leaves. Mosses lack this ability because they lack stomata. The value of $\delta^{13}\text{C}$ in sedges reflects primarily changes in the water budget of the plant and changes in the $\delta^{13}\text{C}$ of atmospheric CO₂ ($\delta^{13}\text{C}_{\text{atm}}$), whereas $\delta^{13}\text{C}$ in mosses also responds strongly to changes in atmospheric concentration of CO₂. By combining $\delta^{13}\text{C}$ measurements in mosses and sedges with estimates of past atmospheric $\delta^{13}\text{C}$, one can reconstruct past CO₂ variations.

We have modelled the $\delta^{13}\text{C}$ values of plants as a function of environmental variables using the set of equations given in Box

BOX 1 Equations governing the $\delta^{13}\text{C}$ values of mosses and sedges as a function of environmental factors

General C3 plants

$$\delta^{13}\text{C}_{\text{plant}} = \delta^{13}\text{C}_{\text{atm}} - a - (b-a) \left(\frac{[\text{CO}_2]_i}{[\text{CO}_2]_a} \right) + \frac{f \Gamma^* + e R_d/k}{[\text{CO}_2]_a} \quad (1)$$

$$A_{\text{plant}} = \frac{[\text{CO}_2]_a - [\text{CO}_2]_i}{r} \quad r = r_a + r_d + r_m \quad (2)$$

$$\delta^{13}\text{C}_{\text{plant}} = \delta^{13}\text{C}_{\text{atm}} - a - (b-a) \left(1 - \frac{rA}{[\text{CO}_2]_a} \right) + \frac{f \Gamma^* + e R_d/k}{[\text{CO}_2]_a} \quad (3)$$

Mosses

$$A_{\text{moss}} = P_{\text{max}} f(I) v(T) i(W) j([\text{CO}_2]_a) k(O) \quad r = \text{constant} \quad (4)$$

$$f(I) = \frac{I - I_c}{I + I_m} \quad (5)$$

$$P_{\text{max}} = P_{\text{max}}^0 (1 + \lambda \Delta T) \quad (6)$$

$$i(W) = \frac{W - W_{\text{dry}}}{W_{\text{opt}} - W_{\text{dry}}} \quad i(W) = 1 \text{ if } W > W_{\text{opt}} \quad i(W) = 0 \text{ if } W < W_{\text{dry}} \quad (7)$$

$$j([\text{CO}_2]_a) = \alpha + \beta [\text{CO}_2]_a \quad (8)$$

Sedges

$$r = \left(g_0 + g_1 \frac{A_{\text{sedge}} h}{[\text{CO}_2]_a} \right)^{-1} \quad (9)$$

$$\Gamma^{13}\text{C}_{\text{sedge}} = \delta^{13}\text{C}_{\text{atm}} - a - (b-a) \left(1 - \frac{1}{h g_1} \right) + \frac{f \Gamma^* + e R_d/k}{[\text{CO}_2]_a} \quad (10)$$

For C3 plants in general, isotopic fractionations (a , b , e and f in above equations) are associated with (1) molecular diffusion in the stomatal recesses for sedges ($a = 4.4\text{‰}$) and turbulent diffusion in the air above the leaf surfaces plus internal effects for mosses (estimated $a = 3\text{‰}$); (2) enzyme-controlled photosynthesis reaction ($b = 30\text{‰}$); (3) dark respiration ($e = 7\text{‰}$ (G. Farquhar, personal communication)); $R_d = 0.06 A_{\text{moss}}$; ($k = (A_{\text{moss}} + R_d/C_i - \Gamma^*)$) and (4) photorespiration ($f = 2\text{‰}$; Γ^* is a function of temperature³¹ with a value of 25 p.p.m. at 10 °C). $[\text{CO}_2]_i$ and $[\text{CO}_2]_a$ are CO₂ concentrations inside the leaf and in the ambient air. A is the assimilation rate of carbon in $\mu\text{mol m}^{-2} \text{s}^{-1}$. The total resistance, r , to CO₂ diffusion is the sum of the atmospheric boundary layer resistance (r_a), the stomatal resistance (r_d) and the mesophyll resistance (r_m). For mosses, r is dominated by r_a (refs 32, 33). Variables controlling the assimilation rate, A_{moss} , are light (photosynthetically active radiation, I in W m^{-2}), temperature (T), water content (W in per cent by dry weight), carbon dioxide concentration ($[\text{CO}_2]_a$ in p.p.m.) and other (O) factors such as nutrients. Compensation point, I_c and half-saturation point, I_m are 5 and 30 W m^{-2} , respectively. P_{max}^0 is the maximum assimilation rate for modern conditions ($T_0 \approx 7^\circ\text{C}$) and ΔT is the assumed temperature variation in the past, scaled by an adaptation factor, λ . Dry and optimal water content W_{dry} and W_{opt} , are 0 and 80%, respectively. The response of A_{moss} to CO₂ is taken from *Dicranum majus* ($\beta = 3.7 \times 10^{-3}$ per p.p.m. between 200 and 400 p.p.m.). The value of α is adjusted to yield pre-industrial $[\text{CO}_2]_a$ of 280 p.p.m. As a point of reference, a 10% decrease in A_{moss} translates into a 3.8‰ decrease in $\delta^{13}\text{C}_{\text{moss}}$. For sedges, r is dominated by r_d which depends on $[\text{CO}_2]_a$, A_{sedge} and relative humidity, h . g_0 and g_1 are constants.

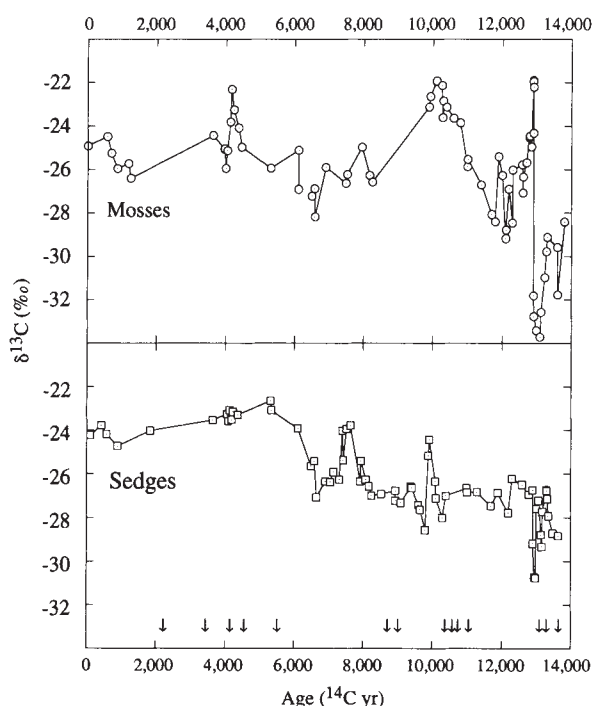


FIG. 1 $\delta^{13}\text{C}$ values of cellulose extracted from mosses and sedges in the Harberton peat core plotted against ^{14}C age. Individual data points are 1-cm increments, representing ~10 yr. The experimental precision in the determination of $\delta^{13}\text{C}$, based on replicate measurements at several depths, is $\pm 0.1\text{‰}$; this is less than the size of the plot symbol. Dominant species between 0 and 6,500 yr BP is *Sphagnum magellanicum* and below that, *Drepanocladus* spp. Ages are interpolated between the Accelerator Mass Spectrometry-dated depths on the core denoted by vertical arrows. ($\delta^{13}\text{C} = ([^{13}\text{C}/^{12}\text{C}]_{\text{sample}} / [^{13}\text{C}/^{12}\text{C}]_{\text{standard}}) - 1$; the PDB standard has been used.)

1. Equation (1), called the discrimination equation, holds for any C3 plant^{7,8}. It accounts for both kinetic and biochemical fractionation of carbon isotopes accompanying the uptake of CO₂ from the ambient air ([CO₂]_a) to the chloroplast ([CO₂]_i). Also included are the fractionations associated with photo-respiration and dark-respiration. The [CO₂]_i/[CO₂]_a ratio is related to the net carbon assimilation rate (A_{plant}) by means of the overall resistance to diffusion (r) as expressed in equation (2). This equation can be solved for the [CO₂]_i/[CO₂]_a ratio, yielding equation (3) in which $\delta^{13}\text{C}_{\text{plant}}$ is related to environmental factors that can be measured or estimated as detailed below.

The assimilation rate of mosses (A_{moss}) is modelled in equation (4) as the product of independent responses to light, temperature, water content, atmospheric carbon dioxide and other factors which may include canopy shading, pH and nutrients. Lacking detailed information for this exercise, we will ignore the effects of these other factors and set $k(0)=1$ in equation (4). Although nutrient availability did change in the past, the presence in the core of abundant charcoal around 9,800 yr BP is not correlated with changes in $\delta^{13}\text{C}_{\text{moss}}$, suggesting little impact of nutrients on A_{moss} . Over the course of a growing season, A_{moss} is generally not light limited⁹. Integrated over a decade, $f(I)$, the impact of changes in light, in equation (5) (ref. 10) is ~ 1 . The impact of extended periods of cloudiness is not clear, however, and remains a potential problem. We calculate that orbitally induced variations in insolation cause $\delta^{13}\text{C}_{\text{moss}}$ to vary by $<0.5\text{‰}$, which is small compared to the variability in this quantity (Fig. 1).

Studies of the effect of temperature on A_{moss} describe a broad maximum at an optimum temperature⁹, although mosses probably acclimatize to the ambient ground temperature with a response time of days to weeks^{11,12}. Considering the time interval integrated in our measurements, we assume mosses to be thermally acclimatized and set $v(T)$, the impact of changes in temperature, equal to 1 in equation (4). On longer timescales, temperature may still affect the maximum rate of photosynthesis, P_{max} , through adaptation. Although such an effect has not been clearly demonstrated and the variation between species may be large¹³, an adaptation to temperature yielding a 5% change in P_{max} per °C (equation (6)) appears reasonable in view of the existing data.

Because mosses can absorb or lose water more readily than vascular plants¹⁴, A_{moss} depends greatly on water availability, generally expressed as W , water content in per cent⁹. We describe the sensitivity of A_{moss} to W according to equation (7) (ref. 12), where W_{dry} and W_{opt} denote respectively the value below which $i(W)=0$, and the optimum value above which $i(W)=1$. We use relative humidities derived from the measurement of $\delta^{13}\text{C}_{\text{sedge}}$ along the core as a proxy for W in the past¹⁵. Neglecting the small respiration terms in equation (1), $\delta^{13}\text{C}_{\text{sedge}}$ reflects primarily relative humidity. This is because the primary resistance of vascular plants to CO₂ uptake and water loss (the stomatal resistance) varies according to equation (9) (refs 16, 17). Provided that the intercept g_0 is small, equation (9) can be combined with equation (3) to yield equation (10) in which $\delta^{13}\text{C}_{\text{sedge}}$ is inversely related to the relative humidity. It can be seen that factors other than humidity, (such as light or temperature) that affect A_{sedge} do not directly affect the value of $\delta^{13}\text{C}_{\text{sedge}}$. We solve equation (10) at each time step to calculate humidities from $\delta^{13}\text{C}_{\text{sedge}}$, the value for g_1 being chosen to yield the modern observed humidity.

Whereas most C3 plants maintain a nearly constant [CO₂]_i/[CO₂]_a ratio¹⁸ in spite of large changes in [CO₂]_a, mosses, lacking stomata, cannot regulate CO₂ uptake or loss. It has been established¹⁹ that A_{moss} is approximately a linear function of [CO₂]_a (equation (8)). Experimental determinations of α and β do not exist for *Sphagnum magellanicum* and *Drepanocladus* in the [CO₂]_a range characteristic of Holocene–Pleistocene changes. By default, we use the value of β determined experimentally for the moss *Dicranum majus*¹⁹.

Using equations (4) to (10), we solve equations (3) to infer [CO₂]_a in the past. The value of $\delta^{13}\text{C}_{\text{atm}}$ can be derived from isotopic measurements on C4 plants²⁰ or from ice cores²¹. The slight difference in the $\delta^{13}\text{C}_{\text{atm}}$ records from these two approaches has only a small effect on the calculated values, but the lack of $\delta^{13}\text{C}_{\text{atm}}$ estimates on the same timescale as our measurements necessitates interpolation, which increases the uncertainty in our reconstructions. In this initial study, we have chosen not to weight the humidities calculated from $\delta^{13}\text{C}_{\text{sedge}}$, or the final [CO₂]_a, by the assimilation rate of carbon. This may exaggerate the importance of periods of low assimilation, for example during times of water stress or low light.

Figure 2 shows the sensitivity of our reconstructed [CO₂]_a to the modelled response to temperature, water content and atmospheric CO₂. In all cases, the product $P_{\text{max}} \times r \times \alpha$ is set by forcing the average, reconstructed pre-industrial [CO₂]_a equal to 280 p.p.m. (refs 22, 23, 24). The largest effect is that of water content; $\pm 30\%$ in W_{opt} yields $\sim \pm 10$ p.p.m. at 10,500 yr. A $\pm 20\%$ change in β translates to about ± 5 p.p.m. effect. Changes in P_{max} with temperature yield only a difference of about ± 5 p.p.m. in the reconstructed CO₂ values.

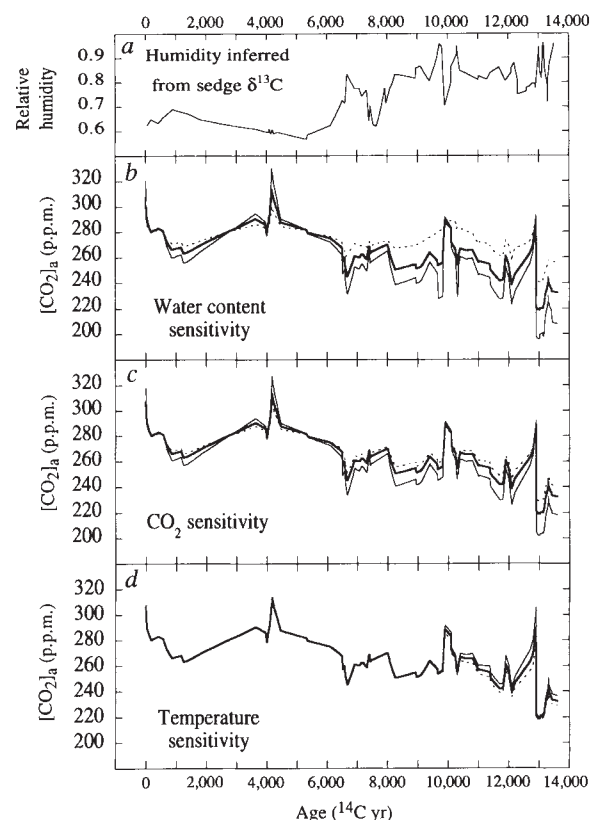


FIG. 2 a, Relative humidity inferred from measurements of $\delta^{13}\text{C}_{\text{sedges}}$. b, c, d, Model reconstructed atmospheric concentration of CO₂, [CO₂]_a, for different responses of A_{moss} to water content, atmospheric CO₂ and temperature. Heavy solid line corresponds to the 'control run' with parameters described in the text, solid line to the increased value and dashed line to the decreased value. b, The optimum water content (W_{opt} in equation (7)) is varied by $\pm 30\%$ about a control run value of 80%. Because of the non-linear response of A_{moss} to W (equation (7)), the dashed and solid lines are not symmetric about the control. c, The sensitivity to CO₂ (β in equation (8)) is varied by $\pm 20\%$ about the control run value (Box 1). The same value of β is used for both *Sphagnum magellanicum* and *Drepanocladus* spp. d, The sensitivity to temperature (λ in equation (6)) is varied by $\pm 1 \mu\text{mol s}^{-1} \text{ °C}^{-1}$ about the control run value of 0. For this sensitivity study, T is arbitrarily set constant between 0–10,000 yr BP and then decreases linearly by 1 °C per 1,000 yr.

FIG. 3 Comparison of $[CO_2]_a$ reconstructed from peat data with measurements from the Byrd ice core, considered to be the best, high resolution record covering this time period. The Byrd dating is from ref. 25, absolute dates being converted to the ^{14}C scale using ref. 34. The parameters used in the model are those of the control run (Fig. 2). The only tuning done to match the two curves is to set the pre-industrial Holocene value of $[CO_2]_a$ to 280 p.p.m. in the model. Although some features, such as the broad hump observed between 1,000 and 7,000 yr BP, may be artefacts of uncertain sensitivity parameters in the model (Fig. 2), the rapid increases in $[CO_2]_a$ at 12,800, 10,000 and 4,400 yr BP are believed to be real.

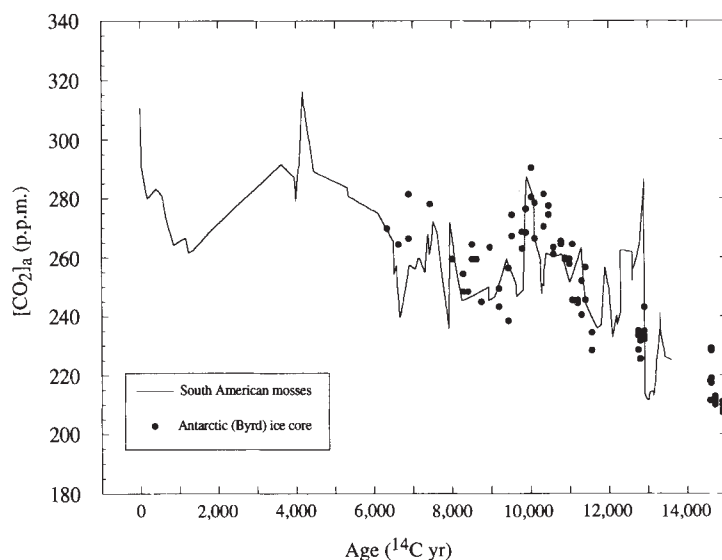


Figure 3 shows good agreement, both for the magnitude of the changes and for the temporal trends, between $[CO_2]_a$ reconstructed from peat data and from measurements from the Byrd ice core²⁵. The occlusion of air bubbles in ice inherently smoothes $[CO_2]_a$ by ~ 70 yr in this ice core (ref. 26); peat data represent $[CO_2]_a$ on a roughly decadal basis. Three prominent features in our reconstructed $[CO_2]_a$ curve cannot be ascribed to an uncertain sensitivity to climate change. The first is a large, rapidly rising spike of ~ 50 p.p.m. in 50 yr at 12,800 yr BP. This increase is roughly the same rate and magnitude as the present day, anthropogenically induced increase. Confirmation of this event is precluded by gaps in the Byrd ice core data. The second feature, which is also observed in the Byrd core, is the large peak at $\sim 10,000$ yr BP. The third feature is a spike beginning at $\sim 4,400$ yr BP. We know of no evidence for or against this spike. All three spikes are associated with major global changes in climate. There is global evidence for the onset of a major warm period at 12,800 yr BP, termed the Bølling warming in northern Europe; 10,000 yr BP marks the termination of the Younger Dryas cold period; and around 4,400 yr BP a global-scale change in climate occurred, and truly modern climates became established, including the onset of El Niño/Southern Oscillation (ENSO)^{27,28}. The Holocene hump in $[CO_2]_a$ on which the 4,400 yr BP feature rests may result from an over-estimated sensi-

tivity of *Sphagnum* to $[CO_2]_a$ in the model. The $[CO_2]_a$ increase from pre-industrial times to the present is driven primarily by the decreasing trend of $\delta^{13}C_{atm}$ prescribed in the model²¹.

Without a high-resolution record of $\delta^{13}C_{atm}$, it is difficult to assess the exact magnitude of the inferred rapid changes in $[CO_2]_a$. If, for example, the 50 p.p.m. rise found at 12,800 yr BP was caused by a massive release of biospheric carbon, one would expect a decrease in $\delta^{13}C_{atm}$, and thus the inferred $[CO_2]_a$ increase would be smaller. This would not be the case, however, if the ocean were responsible for this increase by means of, for example, the suspected re-establishment of deep water formation^{29,30} as the ocean-atmosphere $\delta^{13}C$ fractionation is much smaller.

We describe here a new method for reconstructing $[CO_2]_a$, with high temporal resolution, from the combination of $\delta^{13}C$ measurements in sedges and mosses. Despite the lack of detailed information on the response of mosses to climate, the reconstructed $[CO_2]_a$ curve agrees well with ice-core data, implying that $[CO_2]_a$ may have changed in the past as rapidly as the present, anthropogenically-induced increase. Our method is eminently testable. Any well-preserved deposit where both sedges and mosses are found, should yield the same $[CO_2]_a$ curve, despite different local climate histories. □

Received 25 June; accepted 16 November 1993.

- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
- Markgraf, V. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **102**, 53–68 (1993).
- Markgraf, V., Dodson, J. R., Kershaw, A. P., McGlone, M. S. & Nicholls, N. *Clim. Dynam.* **6**, 193–211 (1991).
- DeNiro, M. J. *Earth planet. Sci. Lett.* **54**, 177–185 (1981).
- Epstein, S., Yapp, C. J. & Hall, J. H. *Earth planet. Sci. Lett.* **30**, 241–251 (1976).
- Green, J. W. in *Methods in Carbohydrate Chemistry* (ed. Whistler, R. L.) 110–135 (Academic, New York, 1963).
- Farquhar, G. D., Hubick, K. T., Condon, A. G. & Richards, R. A. in *Stable Isotopes in Ecological Research* (eds Rundel, P. W., Ehleringer, J. R. & Nagy, K. A.) 21–40 (Springer, New York, 1988).
- Farquhar, G. D., O'Leary, M. H. & Berry, J. A. *Aust. J. Plant Physiol.* **9**, 121–137 (1982).
- Oechel, W. C. & Sveinbjörnsson, B. in *Vegetation and Production Ecology of the Alaskan Arctic Tundra* (ed. Tieszen, L. L.) 269–298 (Springer, New York, 1978).
- Bonan, G. B. *J. geophys. Res.* **96**, 7301–7312 (1991).
- Oechel, W. C. *Photosynthetica* **10**, 447–456 (1987).
- Miller, P. C. et al. *Ecol. Monogr.* **54**, 361–405 (1984).
- Longton, R. E. *The Biology of Polar Bryophytes and Lichens* (Cambridge Univ. Press, 1988).
- Sveinbjörnsson, B. & Oechel, W. C. in *Bryophytes and Lichens in a Changing Environment* (eds Bates, J. W. & Farmer, A. M.) (Oxford Univ. Press, 1992).
- Dilks, T. J. K. & Proctor, M. C. F. *New Phytol.* **82**, 97–114 (1979).
- Reynolds, F. A. et al. *Modeling the Response of Plants and Ecosystems to Elevated CO₂ and Climate Change* (US Dep. of Energy, Washington DC, 1992).
- Ball, T. thesis, Stanford Univ. (1988).
- Polley, H. W., Johnson, H. B., Marino, B. D. & Mayeux, H. S. *Nature* **361**, 61–64 (1993).
- Silvola, J. *Lindbergia* **11**, 86–93 (1985).

- Marino, B. D., McElroy, M. B., Salawitch, R. J. & Spaulding, W. G. *Nature* **357**, 461–466 (1992).
- Leuenberger, M., Siegenthaler, U. & Langway, C. C. *Nature* **357**, 488–490 (1992).
- Neftel, A., Moor, E., Oeschger, H. & Stauffer, B. *Nature* **315**, 45–47 (1985).
- Friedli, H., Löffler, H., Oeschger, H., Siegenthaler, U. & Stauffer, B. *Nature* **324**, 237–238 (1986).
- Siegenthaler, U. & Oeschger, H. *Tellus* **39B**, 140–154 (1987).
- Neftel, A., Oeschger, H., Stauffer, B. & Siegenthaler, U. *Nature* **331**, 609–611 (1988).
- Sowers, T., Bender, M., Raynaud, D. & Korotkevich, Y. S. *J. geophys. Res.* **97**, 15683–15697 (1992).
- McGlone, M. S., Kershaw, A. P. & Markgraf, V. in *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation* (eds Diaz, H. F. & Markgraf, V.) (Cambridge Univ. Press, 1992).
- Huntley, B. & Webb, T. (eds) *Handbook of Vegetation Science* (Kluwer, Dordrecht, 1988).
- Charles, C. & Fairbanks, R. *Nature* **355**, 416–419 (1992).
- Lehman, S. J. & Keigwin, L. D. *Nature* **356**, 757–762 (1992).
- Brooks, A. & Farquhar, G. D. *Planta* **165**, 397–406 (1985).
- Proctor, M. C. F., Raven, J. A. & Rice, S. K. *J. Bryol.* **17**, 193–202 (1992).
- Proctor, M. C. F. in *Plants and their Atmospheric Environment* (eds Grace, J., Ford, E. D. & Jarvis, P. G.) (Blackwell, Oxford, 1981).
- Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **345**, 405–410 (1990).

ACKNOWLEDGEMENTS. We thank T. Sowers and M. Ankin for the chronology and data for the Byrd ice core; J.-M. Barnola for additional ice core data; P. Tans, M. Troller, G. Farquhar and L. Klinger for reviews and comments; and C. Brock and P. Gemery for help with the measurements and for preparation of the manuscript. This research was supported by the US National Science Foundation.

Off-axis volcanism at the East Pacific Rise detected by uranium-series dating of basalts

Steven J. Goldstein*, Michael R. Perfit†, Rodey Batliza‡, Daniel J. Fornari§|| & Michael T. Murrell¶

* Los Alamos National Laboratory, MS K484, Los Alamos, New Mexico 87545, USA

† Department of Geology, University of Florida, Gainesville, Florida 32611, USA

‡ School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

§ Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

¶ Los Alamos National Laboratory, MS J514, Los Alamos, New Mexico 87545, USA

RECENT detailed surveys of the East Pacific Rise have revealed the complexity of the volcanic and magmatic processes occurring along and across fast-spreading ocean ridge crests^{1–7}. In parallel with geological and geochemical investigations, it is now possible to investigate the temporal and spatial pattern of volcanism at ocean ridges by dating young basalts using mass spectrometric uranium-series disequilibria methods^{8–11}. Here we use ²³⁸U–²³⁰Th and ²³⁵U–²³¹Pa ages for basalts to quantify the spatial extent of young volcanism and crustal accretion at 9°31' N on the East Pacific Rise. Most of the ages are younger than would be expected based on off-axis distance and spreading rate. We infer from these anomalously young ages that most of the dated basalts on the crestal plateau were erupted 0.5–2 km outside the axial summit caldera, with some volcanism occurring as far as 4 km off-axis. Melts erupted outside the axial summit caldera can have crustal residence times and magmatic supply systems that differ from those of axial lavas.

The East Pacific Rise (EPR) between the Clipperton and Siqueiros transform zones (Fig. 1) is a well studied, fast-spreading mid-ocean ridge segment (5.5 cm yr⁻¹ half rate)^{12,13}. Much of the segment has an inflated, 4 to 8-km-wide rectangular cross-section^{14–16} with a well-defined axial summit caldera (ASC) 40–300 m wide^{2,6}. A seismic reflector believed to represent the top of a ~1–2-km-wide axial melt lens occurs ~1.5 km below the sea floor^{17–19}. Recent volcanic activity has been documented within the ASC^{2,20}, with the most recent eruption identified at 9°45'–9°52' N in April–May 1991 (ref. 34).

Locations for samples near 9°31' N are shown in Fig. 1, and distances from the ASC and ages are listed in Table 1. Basalts were recovered from the ASC and the adjacent crestal plateau between 9°29' and 9°33' N using the submersible *Alvin*, a precision-navigated rock corer^{5–7} and a conventional dredge²¹. Nine *Alvin* samples from the ASC between 9°17' and 9°54' N were also analysed to determine the initial U–Th–Pa systematics for recent (<1 kyr)² volcanism within the ASC.

²³⁰Th and ²³¹Pa model ages were determined by estimating initial ²³⁰Th/²³²Th and ²³¹Pa/²³⁵U ratios from young samples located within the ASC, and by assuming constancy of these initial ratios over time. Initial ²³⁰Th/²³²Th ratios, and to a lesser extent initial ²³¹Pa/²³⁵U ratios, are dependent on incompatible element ratios and were estimated from trends of ²³¹Pa/²³⁵U and ²³⁰Th/²³²Th against Th/U for samples within or near the ASC. We also assume that the melting processes creating U-series disequilibria have been uniform for the 9–10° N segment over the

past 130 kyr (ref. 10). That these assumptions are justified is indicated by the general concordance of ²³¹Pa, ²³⁰Th and ²²⁶Ra ages for individual basalt samples from the 9–10° N EPR and Juan de Fuca/Gorda ridges^{9–11}. The accuracy of the ages is also supported by the agreement of spreading rates based on ²³⁰Th ages, palaeomagnetic ages and morphologic features of samples located outside the neovolcanic zone of the Juan de Fuca and Gorda ridges⁹. A more detailed discussion of the data and procedures used to determine ²³⁰Th and ²³¹Pa model ages is presented in ref. 10.

Model ages represent age differences for samples since U–Th–Pa fractionation during partial melting. Consequently the model age T is given by $T = T_t + T_{mc} + T_c$, where T_t is the transit time for melts from mantle to crust, T_{mc} is the residence time within a crustal magma chamber and T_c is the period of time since the eruption. Assuming that transit times are small^{11,22} or constant for samples from similar locations and sources (for example, similar Th/U ratios), the model age can be approximated by the crustal residence age T_{cr} , where $T \approx T_{cr} = T_{mc} + T_c$. Hence, the crustal residence age provides an upper limit on the eruption age, and $T_c < T_{cr}$. But for older samples located far off-axis, where magma chamber ages are small relative to eruption ages ($T_{mc} \ll T_c$), then $T_c \approx T_{cr}$.

Table 1 lists ²³¹Pa and ²³⁰Th model ages and locations for samples collected outside the ASC near 9°31' N on the EPR. Ages are plotted against distance from the ASC in Fig. 2a. Model ages for seven of the nine normal MORB (N-MORB) samples are younger (by up to 70 kyr) than predicted from their palaeomagnetic ages, whereas the other N-MORB samples have crustal ages approximately equal to their palaeomagnetic ages. Anomalously young N-MORB samples are found up to 3.8 km from the ASC on the Cocos plate, and 2.5 km from the ASC on the Pacific plate. Based on the age anomalies, most of the dated N-MORB samples were erupted 0.5–2.2 km outside the ASC, with two samples erupted within the ASC and one sample emplaced ~4 km outside the ASC. These data indicate that the neovolcanic zone is at least 4–8 km wide, and can extend throughout the crestal plateau of the EPR.

Two enriched MORB (E-MORB) samples (A2489-3, ARC 48) located 2.3–2.5 km east of the ASC have crustal residence ages which are 10–20 kyr older than palaeomagnetic ages based on spreading rate, whereas the other E-MORB sample (R49-4) has a younger crustal age. The greater ages for the two

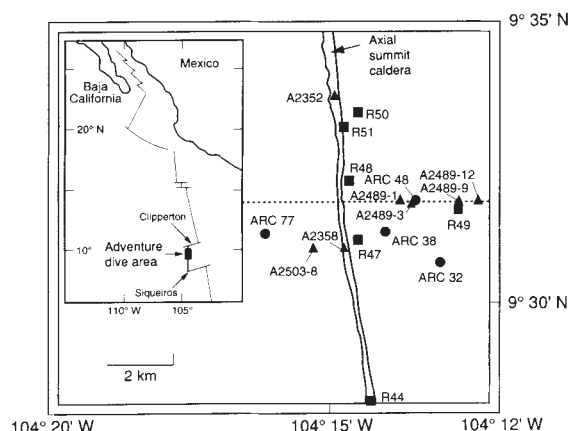


FIG. 1 Map of 9°31' N area illustrating locations of samples collected by submersible (triangles), rock core (circles) and dredge (squares), and the axial summit caldera (ASC)^{2,6}. Dotted line shows the approximate track of *Alvin* traverses of the crestal plateau during dives 2,488, 2,489 and 2,495. Dredges R54, R57 and R58 are off the northern edge of the figure, between 9°35' and 9°40' N, and dredge R39 is at 9°21' N. Distances from ASC for additional samples recovered from within and near the ASC by *Alvin* and dredge are listed in Table 1.

|| Present address: Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

E-MORB lavas could indicate that they had a significantly longer magma chamber residence period than the near-axis E-MORB sample. Alternatively, the greater ages could reflect asymmetric spreading, with slower spreading for the Cocos plate.

One sample (ARC-38) located ~1.3 km east of the ASC has a lower ^{231}Pa age (0 kyr) than two samples from within the ASC at the same latitude (A2358-3 and -4; 4–5 kyr). Hence in at least one instance, magma supplying eruptions outside the ASC has

either experienced a shorter crustal residence than magma supplying the nearby ASC, or was derived from a source with slightly different initial $^{231}\text{Pa}/^{235}\text{U}$ ratio than axial samples. The similar Th/U ratio for ARC-38 compared to the ASC lavas¹⁰ lends support to the first hypothesis. In either case, the data suggest that eruptions outside the ASC are fed by crustal plumbing systems that can be separated from those supplying volcanism within the ASC, and that melts erupted on the crestal plateaus of fast-spreading ocean ridges can bypass the central portion of the sub-crestal magma lens.

Figure 2b shows previously published ^{230}Th ages for the Endeavour segment of the Juan de Fuca ridge⁹. Locations of the zone of focused axial volcanism and off-axis samples at Endeavour are less precisely known than at $9^\circ 31' \text{N}$ EPR, and ^{230}Th ages for Endeavour are less extensive, precise and conclusive than the concordant ^{231}Pa and ^{230}Th ages for the EPR. However, most of the off-axis Endeavour lavas are also younger than predicted (based on palaeomagnetic ages) and suggest a ~4 to 6-km-wide neovolcanic zone. Thus, the available U-series age data indicate that volcanism can occur throughout the crestal region of both intermediate- and fast-spreading ocean ridges.

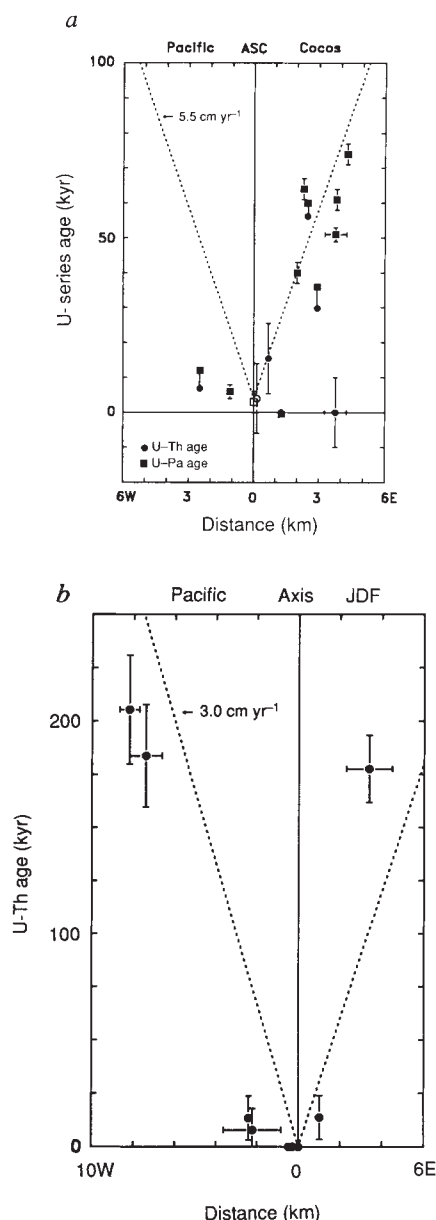


FIG. 2 U-series age versus distance off-axis for a, $9^\circ 31' \text{N}$ EPR (data from Table 1) and b, the Endeavour segment of Juan de Fuca ridge (data from ref. 9). Dashed lines indicated palaeomagnetic half-spreading rates, whereas error bars illustrate start/end positions for dredge samples and estimated 2σ uncertainties in ^{230}Th and ^{231}Pa ages. For simplicity, only the average and range of data for EPR $9^\circ 31' \text{N}$ samples located within or near the ASC are plotted as open symbols. Data points for samples with both ^{230}Th and ^{231}Pa ages are connected by tie lines, and error bars are omitted. For the EPR, anomalously young ages for most samples outside the ASC indicate a neovolcanic zone ~4–8 km wide. Results for the intermediate-spreading Endeavour ridge also show anomalously young ages and indicate a neovolcanic zone ~4–6 km wide. For both the eastern flanks of the EPR and Endeavour ridge, samples with ages older than palaeomagnetic ages may reflect asymmetric spreading, with slower spreading to the east⁹.

TABLE 1 Locations and ages of the $9\text{--}10^\circ \text{N}$ EPR samples

Sample	Distance from ASC* (km)	Spreading age† (kyr)	^{230}Th age‡ (kyr)	^{231}Pa age‡ (kyr)	Site of eruptions§ (km off-axis)
<i>Alvin</i>					
A2352-2	0.0	0	2 ± 9	3 ± 2	—
A2355-8	0.0	0	0 ± 8	3 ± 2	—
A2356-7	0.0	0	0 ± 8	1 ± 2	—
A2358-3	0.0	0	—	4 ± 2	—
A2358-4	0.0	0	—	5 ± 2	—
A2359-4	0.0	0	—	2 ± 2	—
A2359-5	0.0	0	—	3 ± 2	—
A2365-3	0.0	0	—	3 ± 2	—
A2372-1	0.0	0	—	0 ± 2	—
A2489-1	2.0 E	36	—	40 ± 3	0.0
A2489-3	2.3 E	42	—	64 ± 3	—
A2489-9	3.8 E	68	—	61 ± 3	0.6 E
A2489-12	4.3 E	78	—	74 ± 3	0.4 E
A2503-8	1.1 W	20	—	6 ± 2	1.0 W
Dredge samples (near ASC)					
R39-1	0.0	0	1 ± 11	—	—
R44-6	0.0	0	14 ± 9	—	—
R47-3	0.4 E	7	5 ± 9	—	—
R48-1	0.3 E	5	5 ± 9	—	—
R51-1	0.1 E	2	0	—	—
R54-2	0.4 E	7	0	0	—
R57-6	0.3 E	5	17 ± 11	—	—
R58-1	0.4 E	7	7 ± 15	—	—
Rock core and dredge samples (outside ASC)					
ARC-32	2.9 E	53	29 ± 9	36 ± 3	1.4 E
ARC-38	1.3 E	24	0 ± 12	-1 ± 4	1.3 E
ARC-48	2.5 E	45	56 ± 9	60 ± 3	—
ARC-77	2.5 W	45	7 ± 8	12 ± 3	2.2 W
R49-1	3.8 E	68	0 ± 11	—	3.8 E
R49-4	3.8 E	68	—	51 ± 2	1.2 E
R50-7	0.7 E	12	15 ± 10	—	0.1 E

* Distances from ASC have uncertainties of less than ± 0.1 km, except for dredge R49 (± 0.5 km).

† Spreading ages based on distance from ASC and a palaeomagnetic half-spreading rate of 5.5 cm yr^{-1} .

‡ Data and ages from ref. 10, except for *Alvin* samples outside the ASC. $^{231}\text{Pa}/^{235}\text{U}$ activity ratios: for 2489-1, 1.738 ± 0.022 ; 2489-3, 1.484 ± 0.014 ; 2489-9, 1.471 ± 0.019 ; 2489-12, 1.356 ± 0.020 ; 2503-8, 2.515 ± 0.021 . Initial $^{231}\text{Pa}/^{235}\text{U}$ and $^{230}\text{Th}/^{232}\text{Th}$ activity ratios and equations used to determine ^{231}Pa and ^{230}Th ages for N-MORB and E-MORB samples (2489-3, ARC48, R49-4) are given in refs 9, 10. Based on concordant ^{230}Th and ^{231}Pa ages, uncertainties in initial $^{231}\text{Pa}/^{235}\text{U}$ and $^{230}\text{Th}/^{232}\text{Th}$ activity ratios over time are based on the observed range for axial samples¹⁰. Errors in ^{230}Th and ^{231}Pa ages are 2σ .

§ Location of eruption for anomalously young samples calculated from difference between (^{230}Th and/or ^{231}Pa age – 4 kyr) and spreading age, assuming a palaeomagnetic half-spreading rate of 5.5 cm yr^{-1} , and where 4 kyr is the average of ^{230}Th and ^{231}Pa ages for samples located near or within the ASC. Samples A2489-3 and ARC-48 yield anomalously old ages; see text for possible explanations.

|| *Alvin* samples were recovered from both within and outside the ASC, near-ASC dredge samples are located < 0.5 km away from the ASC, and other dredge and rock core samples are located > 0.5 km away from the edges of the ASC.

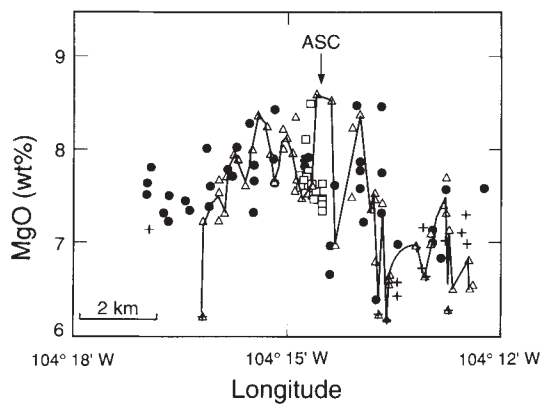


FIG. 3 Across-axis variability of lava MgO content from the 9°29'–9°33' N region of the EPR. Lavas from the axial summit caldera (ASC, open squares) show limited chemical variability compared to those from the crestal plateau. N-MORB samples recovered by *Alvin* from an E–W transect (open triangles; M. R. Perfit and D. J. Fornari, unpublished data) and by rock coring (filled circles, ref. 7) show small-scale compositional variability and asymmetry. E-MORB samples (crosses) are also asymmetrically distributed and only found outside the ASC. Across-axis variations in MgO and other chemical parameters⁷, and the evidence for anomalously young lavas on the crestal plateau, suggest that melts supplying volcanism outside the ASC can have distinct and spatially restricted magmatic supply systems and crustal residence times relative to axial lavas.

The U-series ages provide a direct, quantitative measure of the width of the neovolcanic zone which is broadly consistent with more qualitative or indirect geological and geophysical constraints on crustal accretion and volcanic processes at intermediate- and fast-spreading ocean ridges. The generally rectangular or dome-shaped cross-sectional morphology of the EPR between 9° and 10° N can be explained by coeval and widespread constructional volcanism and intrusion throughout the crestal plateau region, in addition to the thermal uplift caused by abundant magma supply²³. Also, thickening of seismic layer 2A from ~100–200 m to 400–500 m within 1–2 km of the ridge axis at 9°17' N (ref. 24) and from 160 to 340 m within ~1 km of the ridge axis at 9°30' N (ref. 25) is a likely consequence of intrusive and extrusive magmatic activity within ~1–2 km of the axis. Transition widths between opposed magnetic field directions at fast- and intermediate-spreading ridge crests average up to 6 km and are also consistent with the occurrence of volcanism and crustal accretion outside the ASC^{26,27}. Palaeomagnetic studies of basalts from deep-sea drill cores, which show frequent anomalous inclinations and transitions, also favour a diffuse transition width (see, for example, ref. 28). Other observations at mid-ocean ridges, including considerations of the geometry of melt supply^{29–31} and the existence of off-axis ridge-crest hydrothermal activity (see, for example, refs 32 and 33), are also consistent with a broader, 4 to 8-km-wide crustal accretion zone as inferred from the U-series ages.

Recent observations (by submersible) of sediment thickness and volcanic morphology, and geochemical data for lavas outside the ASC at 9°29'–9°32' N on the EPR (Fig. 3; ref. 7) also support the hypothesis that recent volcanism is not confined to the ASC. Sediment cover is anomalously low in areas up to 4 km outside the ASC on the Cocos plate, and 3 km from the ASC

on the Pacific plate. Beyond 3–4 km from the ASC, sediment thickness increases dramatically, and lavas appear to be significantly older, with apparent ages more typical of those predicted on the basis of palaeomagnetic spreading rates (>70 kyr). Although lava that originates in the ASC and flows out through tube and channel systems within the upper volcanic carapace onto the crestal plateau has been observed in some areas along the ASC (for example, at 9°45'–9°54' N), no such break-outs were observed in the 9°31' N region²⁰. Constructional pillow mounds associated with ridge-parallel fissures, and sheet flows with morphologies distinct from those that lie within the ASC⁷ provide additional evidence that these lavas were not emplaced by surficial distributary systems that originated in the ASC. The asymmetric distribution and complex spatial variability of lava compositions throughout the crestal plateau (Fig. 3) are also consistent with the existence of small magma bodies outside the ASC⁷.

We conclude that volcanism outside the ASC is common, and that the magmatic zone is broad (~4–8 km full-width)—certainly broader than the present width of the ASC (~300 m) in the 9°31' N region of the EPR. It is difficult to estimate the extent or volume of volcanism outside of the ASC, but we note that eight out of twelve samples on the crestal plateau are younger than would be predicted from their locations. In addition, the geochronological and geochemical data indicate that lavas on the crestal plateau need not have been erupted through the same crustal plumbing network that supplies melts to the ASC. The common occurrence of anomalously young and compositionally distinct lavas throughout the crestal plateau suggests that melts supplying volcanic centres outside the ASC can have distinct parental magmas, supply systems and crustal residence times relative to axial lavas. □

Received 10 May; accepted 22 November 1993.

- Hekinian, R., Thompson, G. & Bideau, D. *J. geophys. Res.* **94**, 17437–17463 (1989).
- Haymon, R. M. et al. *Earth planet. Sci. Lett.* **104**, 513–534 (1991).
- Sinton, J. M., Smaglik, S. M. & Mahoney, J. J. *J. geophys. Res.* **96**, 6133–6155 (1991).
- Reynolds, J. R. et al. *Nature* **359**, 493–499 (1992).
- Bender, J., Langmuir, C., Reynolds, J., Ryan, W. & Kastens, K. *Ridge Events* **3**, 19–21 (1992).
- Fornari, D. J., Perfit, M., Haymon, R., Edwards, M. H. & O'Brien, T. *Mar. geophys. Res.* (submitted).
- Perfit, M. R. et al. *Geology* (in the press).
- Goldstein, S. J., Murrell, M. T. & Janecky, D. R. *Earth planet. Sci. Lett.* **96**, 134–146 (1989).
- Goldstein, S. J., Murrell, M. T., Janecky, D. R., Delaney, J. R. & Clague, D. A. *Earth planet. Sci. Lett.* **107**, 25–41 (1991); erratum, **109**, 255–272 (1992).
- Goldstein, S. J., Murrell, M. T. & Williams, R. W. *Earth planet. Sci. Lett.* **115**, 151–159 (1993).
- Volpe, A. M. & Goldstein, S. J. *Geochim. cosmochim. Acta* **57**, 1233–1242 (1993).
- Klitgord, K. D. & Mammerickx, J. *J. geophys. Res.* **87**, 6725–6750 (1982).
- Carbotte, S. & Macdonald, K. J. *geophys. Res.* **97**, 6959–6982 (1992).
- Lonsdale, P. *Mar. geophys. Res.* **3**, 295–315 (1977).
- Macdonald, K. C., Sempere, J.-C. & Fox, P. J. *J. geophys. Res.* **89**, 6049–6069 (1984).
- Macdonald, K. C. & Fox, P. J. *Earth planet. Sci. Lett.* **88**, 119–131 (1988).
- Detrick, R. S. et al. *Nature* **326**, 35–41 (1987).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. *Nature* **344**, 650–653 (1990).

- Sinton, J. M. & Detrick, R. S. *J. geophys. Res.* **97**, 6133–6155 (1991).
- Haymon, R. M. et al. *Earth planet. Sci. Lett.* **119**, 85–101 (1993).
- Batiza, R. & Niu, Y. *J. geophys. Res.* **97**, 6779–6797 (1992).
- Rubin, K. H. & Macdougall, J. D. *Nature* **335**, 158–161 (1988).
- Scheirer, D. S. & Macdonald, K. C. *J. geophys. Res.* **98**, 7871–7885 (1993).
- Harding, A., Kent, G., Kappas, M. & Orcutt, J. *Ridge Events* **2**, 8–10 (1991).
- Cristeson, G. L., Purdy, G. M. & Fryer, G. M. *Geophys. Res. Lett.* **19**, 1045–1048 (1992).
- Klitgord, K. D., Huestis, S. P., Mudie, J. D. & Parker, R. L. *Geophys. J. R. astr. Soc.* **43**, 387–424 (1975).
- Macdonald, K. C., Miller, S. P., Luyendyk, B. P., Atwater, T. M. & Shure, L. J. *geophys. Res.* **88**, 3403–3418 (1983).
- Johnson, H. P. & Merrill, R. T. *Earth planet. Sci. Lett.* **40**, 263–269 (1978).
- Spiegelman, M. & McKenzie, D. *Earth planet. Sci. Lett.* **83**, 137–152 (1987).
- Sparks, D. W. & Parmentier, E. M. *Earth planet. Sci. Lett.* **105**, 368–377 (1991).
- Wilson, D. S. *Earth planet. Sci. Lett.* **113**, 41–55 (1992).
- Mottl, M. J. *Geol. Soc. Am. Bull.* **94**, 161–180 (1983).
- Morton, J. L. & Sleep, N. H. *J. geophys. Res.* **90**, 11345–11353 (1985).
- Rubin, K. H., Macdougall, J. D. & Perfit, M. R. *Nature* (submitted).

ACKNOWLEDGEMENTS. This research was supported by the US Department of Energy, Office of Basic Energy Sciences (M.T.M.), and the US National Science Foundation-ODP, MG&G and RIDGE (M.R.P., R.B. and D.J.F.).

Non-universal scaling of fracture length and opening displacement

C. G. Hatton, I. G. Main & P. G. Meredith*

Department of Geology and Geophysics, University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, UK

* Department of Geological Sciences, University College London, Gower Street, London WC1E 6BT, UK

THE deformation of the Earth's brittle crust is dominated by the formation and growth of faults in response to tectonic loading. The scaling properties of such systems provide clues to the underlying mechanisms of fault propagation. For example, if fault growth were a self-similar process, described by a scaling law that applies in all locations¹, this might imply a universal faulting mechanism governed by either constant fracture toughness or constant yield stress. Universal scaling laws have been proposed¹⁻⁴, but their general applicability remains the subject of some debate⁵. In the natural environment, strict scale invariance can apply only between well-defined bounds⁶, and it is known that the Earth's crust has many distinct length scales—ranging from the grain size of rocks and the thickness of sedimentary layers up to the finite width of the seismogenic crust—each of which may vary from place to place. Here we report the scaling properties of two populations of tensile fractures in the Krafla fissure swarm of northeast Iceland (spanning nearly four orders of magnitude in length and five in displacement), which clearly show that the presence of such length scales dramatically alters the scaling behaviour. Despite the geological homogeneity of the region studied, our data cannot be described by a single scaling law.

The Krafla fissure swarm (Fig. 1) is one of five volcanotectonic systems in the active rift zone of northeast Iceland, striking (at 10° E) sub-perpendicular to the spreading direction of the Mid-Atlantic Ridge^{7,8}. The swarm is associated with the Krafla central volcano, and takes the form of an elongate zone (4–10 km wide) of tensile fractures, fissures and normal faults, formed in discrete episodes during the Holocene. It is the most recently active swarm in the northern rift zone, with volcano-tectonic activity beginning in December 1975, and continuing with some activity into 1989. The activity is due primarily to the long-term accumulation of extensional tectonic stress built up from plate movements at the ridge crest, but in this case was triggered by an increase in fluid pressure due to an inflow of magma into the Krafla magma chamber^{9,10}. Ground observations showed that rifting within the fissure swarm involved both the opening of new ground fissures and the reactivation of existing ones^{9,11}.

During a single field season in 1991, a detailed study was made of two populations of tensile fractures in the fissure swarm. In the Kelduhverfi area in the north of the fissure zone the dominant 'fracture' is a normal fault of length 2,275 m with a downthrow to the west varying between 0 and 11 m. Its maximum width, representing the tensile component of opening exposed at the surface, was found to be 18 m. A total of 79 subsidiary fractures were observed in the adjacent relatively homogeneous basaltic lava flows, ranging in length from 0.2 to 245 m. Figure 2 shows the main fault and two examples of smaller-scale associated fractures. The number of observed subsidiary fractures decreased to zero within ~20 m on each side of the dominant fracture, thereby defining the lateral scale of observation. Near Mývatn, to the south of the fissure swarm (Fig. 1), the dominant fissure locally named Grjótagjá was 900 m in length. The associated fractures, in this case within 10 m of this feature, varied in length from 0.4 to 14.8 m, and are significantly shorter than those of the first population of Kelduhverfi.

The length and width of each open fracture in both populations was measured, using the following consistent operational

rules; (1) length defined by a straight line from tip to tip using a tape measure; (2) width defined by the maximum opening displacement along the length of the fracture; (3) joints selected as follows; of the joints with some opening displacement, only those showing evidence of recent opening, with relatively unweathered surfaces and no significant infill of sedimentation or vegetation, were included in the population of measured fractures¹². Columnar cooling joints were found in almost all the lava flows exposed at the surface, with a characteristic exposed dimension (in horizontal section) of ~30 cm. Closed joints were distinguished easily from the relatively fresh exposure of the open tensile fractures, which frequently exploited the weakness of the columnar jointing (for example, Fig. 2c).

The relation between the measured fracture length (L) and maximum opening displacement (D) of the fracture populations

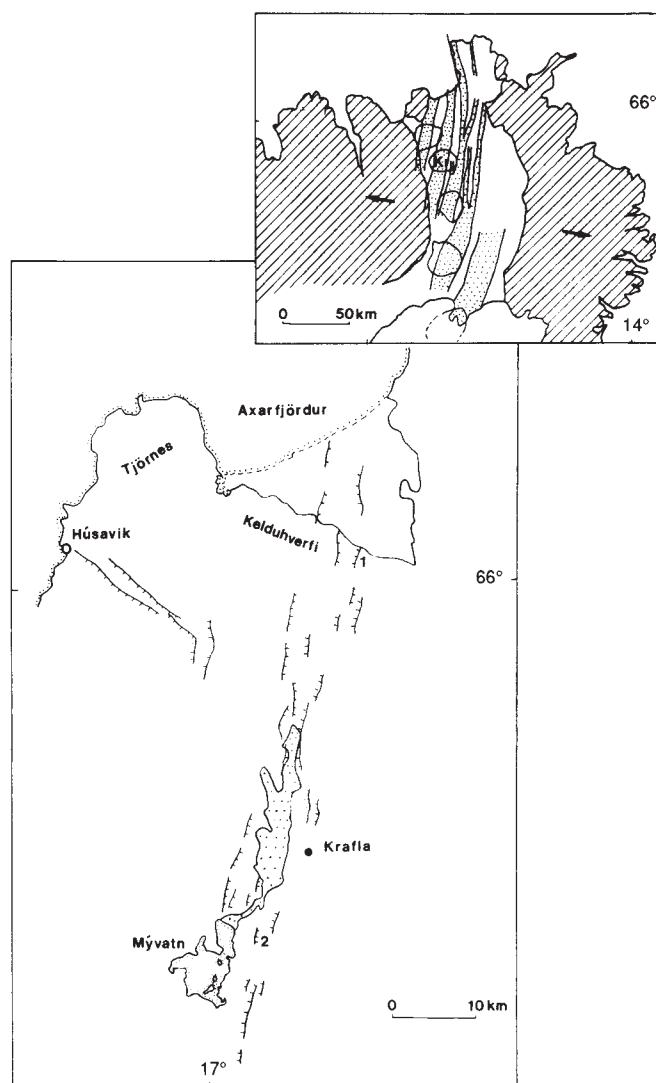
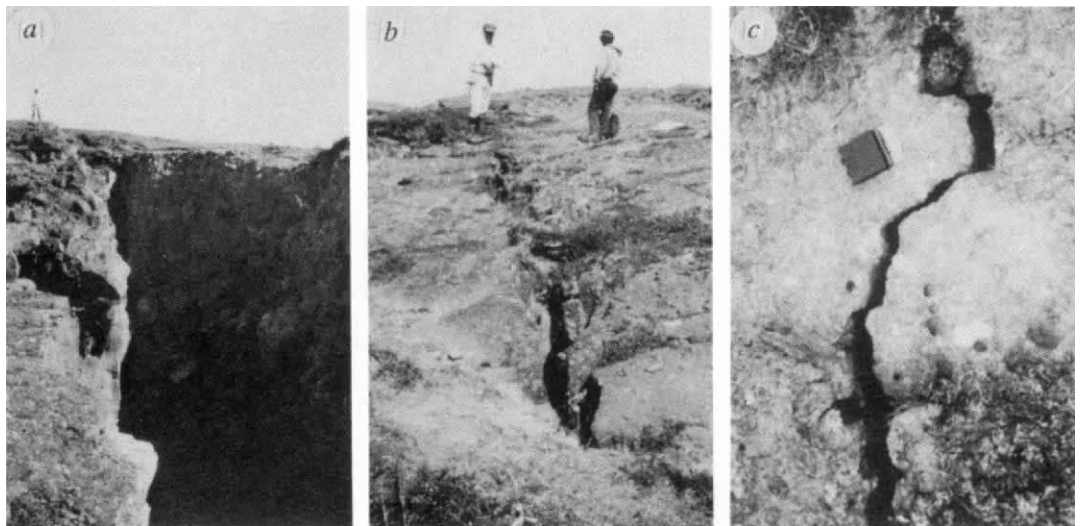


FIG. 1 Inset, A map of northeast Iceland (after Björnsson⁸) showing the orientation of the plate-tectonic spreading direction (arrows), and the trend of the axial rift zone in the centre. Five individual volcano-tectonic fissure swarms are shown (dotted areas), with the Krafla fissure swarm marked Kr. Areas of lava older than 0.7 Myr are indicated by the hatching to either side of the fissure swarms. Main diagram, A simplified map of the Krafla fissure swarm (after Johannesson and Saemundsson²³, with hatched lines indicating exposed normal faults, and areas shaded with dots marking relatively recent lava flows (<1,100 yr old). The localities of two of the fracture populations are marked 1 (Kelduhverfi) and 2 (Mývatn).

FIG. 2 a, The dominant fault in the Kelduhverfi area (looking north), with an opening displacement (width) of 18 m. b, A relatively large intermediate-scale fracture of measured length 21 m. c, A smaller-scale tensile fracture of measured length 1.3 m.



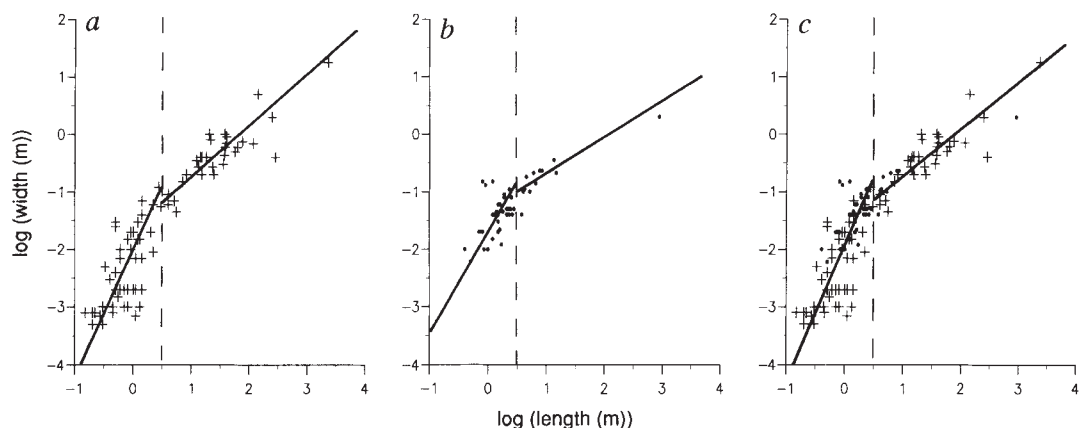
is plotted on log-log axes in Fig. 3. The data from the Kelduhverfi area (Fig. 3a) are not consistent with a universal scaling law, and plot with a distinct break in slope at a length of ~ 3 m. At fracture lengths < 3 m, the data are consistent with a power-law distribution ($D \propto L^n$) with $n = 2.20 \pm 0.24$ (correlation coefficient $r = 0.64$), whereas $n = 0.89 \pm 0.08$ ($r = 0.87$) for fractures above this characteristic length, the errors being expressed as one standard deviation (s.d.). For the Mývatn population (Fig. 3b) the change of scaling at 3 m is less marked, but the data show a change of slope from 1.78 ± 0.47 below 3 m ($r = 0.47$) to 0.63 ± 0.08 ($r = 0.90$) above. For the combined data set (Fig. 3c) the respective slopes are 2.38 ± 0.18 ($r = 0.73$) below 3 m and 0.81 ± 0.06 ($r = 0.88$) above, and there is a strong overlap of the two component populations. This indicates a distinct change in scaling at the 90% confidence level (that is, no overlap in the value of n at ~ 2 s.d.) from $n \approx 2$ to $n \approx 1$ at a fracture length of ~ 10 times the mechanical grain size provided by the cooling joints.

This result suggests that the way in which a fracture grows may not be universal, even during a relatively short period of geological time (several thousand years) in a geologically homogeneous area. That is, under certain circumstances, the growth rule can change once the fracture reaches a critical length, in this case ~ 10 joint spacings. The observation that resistance to crack growth is dependent on crack length, up to a critical value, has been made in the laboratory for many different polycrystalline ceramics^{13,14}, as well as rocks^{15,16}. The critical length corresponds to that of a fully developed cohesive zone, where at least part of the crack is open and hence traction free. Propagation

is then controlled by the bulk properties of the material rather than the local microstructure. For polycrystalline ceramics this change in the growth rule is well documented¹⁷, and is manifested by a sharp increase in the fracture toughness for cracks between one and ten times the crystalline grain size. Unfortunately no published data is available on the scaling properties of crack length and width in ceramics for direct comparison with the present work.

How might the change in scaling at 3 m be controlled by columnar jointing on a scale of 30 cm? One reason might be the effective blunting of the crack tip by the change in mechanical properties as it propagates from one columnar joint to another, illustrated schematically in Fig. 4. Having been arrested by a particular joint oblique to the average direction of fracture propagation, the fracture may accumulate width more easily than length in future discrete episodes of displacement. Some crack blunting (in the form of a cohesive zone of residual frictional contact behind—and a zone of damage ahead of—the crack tip) is a feature of the crack growth model suggested by Barenblatt¹⁸, and applied to fault growth by Cowie and Scholz¹⁹. In our case no examples of residual frictional contact were found for these tensile cracks, but zones of damage in the form of subsidiary microcracking or crack branching were frequently found at the termination of individual cracks. Figure 4 is a schematic illustration of this model applied to the present case. Note that the stress is reduced, compared to the elastic solution, both in the cohesive zone and beyond. If the cohesive zone has a length s , and all other parameters of interest remain constant, then this model predicts that the fault will grow by an amount

FIG. 3 Scaling of fracture width (maximum opening displacement) with length for the fracture populations at: a, Kelduhverfi (crosses); b, Mývatn (dots); and c, a composite plot of the two data sets. The solid lines are least-squares fits to the data above and below 3 m respectively, the divide being shown by the vertical dashed line. The small offset of the two lines at 3 m is not statistically significant.



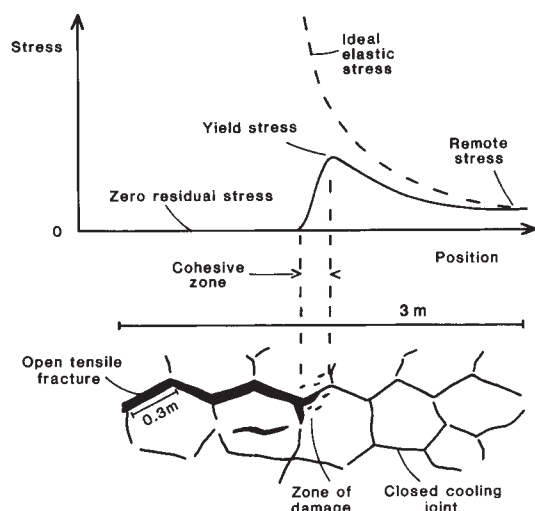


FIG. 4 Bottom, A schematic diagram (plan view) illustrating the blunting of the crack tip due to the influence of the joint pattern, that is forking of the crack tip or the development of a zone of subsidiary damage. Top, A schematic illustration of a model for the anelastic growth of a crack with a cohesive zone^{18,19}. The linear elastic solution is shown by a dashed line for comparison, and leads to a physically unrealistic stress singularity at the crack tip.

$\Delta L/L \propto \sqrt{s/L}$ (ref. 19). Three possibilities then follow from a geometrical fault growth model: (1) $n=2$ when $\Delta L = \text{constant}$, implying a constant growth increment²; (2) $n=1.5$ when $s = \text{constant}$ or $\Delta L \propto \sqrt{L}$, implying a constant fracture toughness²⁰; and (3) $n=1$ when $s \propto L$ or $\Delta L \propto L$, implying a constant yield stress and strict scale-invariance¹⁹. Here the data are consistent with the constant terms in (1) or (2) being determined by the joint spacing, with self-similar growth being allowed only when the crack grows large enough for the resulting increase in stress concentration ahead of the crack tip to allow significant stress transfer to distances larger than a single 'grain' scale. Thus the cooling joints may have acted as a local stress shield for the smaller cracks in either population. Finally, although we are dealing with tensile (mode I) fractures, these results do have significance for the growth models of shear faulting (modes II and III) because the scaling properties predicted by these models are independent of the mode of crack growth, although the absolute values of fracture toughness or yield stress may differ¹⁸.

The average trend of our data sets seems to fit between the two extremes of the models proposed for explaining the growth of faults^{2,19}, but the length-displacement relation is not universal, with a distinct break in scaling from $D \propto L^2$ below 3 m, to $D \propto L$ above. Strictly fractal scaling applies only in the scale range from 3 m to the maximum observed length of 2 km, and probably beyond²¹. Similar breaks in slope might be expected at even greater scales, for example because of the finite width of the seismogenic crust²². Applying a universal average to the whole data set might therefore result in significant systematic errors in calculating the total strain, especially when based on extrapolations, or from composite studies with no strong overlap in scale. Our study represents perhaps one of the simplest cases for investigation. This work also highlights both the dangers of applying universal scaling laws to all tectonic provinces on a range of scales, and the urgent need for further field measurements (over a wide range of scales) of individual populations both of shear faults and tensile fractures. □

3. Marret, R. & Allmendinger, R. W. *J. struct. Geol.* **13**, 735–738 (1991).
4. Walsh, J. J., Watterson, J. & Yielding, G. *Nature* **351**, 391–393 (1991).
5. King, G. & Cisternas, A. *Nature* **351**, 350 (1991).
6. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
7. Saemundsson, K. in *Crustal Evolution in Northwest Britain and Adjacent Regions* (eds Bowes, D. R. & Leake, B. E.) 415–432 (Steel House, Liverpool, 1978).
8. Björnsson, A. *J. geophys. Res.* **90**, 10151–10162 (1985).
9. Tryggvason, E. *Bull. volcan.* **47**, 47–69 (1984).
10. Einarsson, P. *Tectonophysics* **189**, 261–279 (1991).
11. Ophelm, J. A. & Gudmundsson, A. *Geol. Soc. Am. Bull.* **101**, 1608–1622 (1989).
12. Hatton, C. G. thesis, Univ. Edinburgh (1992).
13. Hübner, H. & Jilke, W. *J. Mater. Sci.* **12**, 117–125 (1977).
14. Cook, R. F., Lawn, B. R. & Fairbanks, C. J. *J. Am. ceram. Soc.* **68**, 604–615 (1985).
15. Schmidt, R. A. & Lutz, T. J. in *Fracture Mechanics Applied to Brittle Materials* 166–182 (Spec. Publ. STP678, Am. Soc. Test. Mater., 1979).
16. Ingraffea, A. R. in *Fracture Mechanics of Rock* (ed. Atkinson, B. K.) 71–110 (Academic, London, 1987).
17. Rice, R. W., Freiman, S. W. & Mecholsky, J. J. *J. Am. ceram. Soc.* **63**, 129–136 (1980).
18. Barenblatt, G. I. *Adv. appl. Mech.* **7**, 55 (1962).
19. Cowie, P. A. & Scholz, C. H. *J. geophys. Res.* **97**, 11085–11095 (1992).
20. Gilliespie, P. A., Walsh, J. J. & Watterson, J. *J. struct. Geol.* **14**, 1157–1172 (1992).
21. Gudmundsson, A. *Terra Nova* **4**, 464–471 (1992).
22. Cowie, P. A. & Scholz, C. H. *J. struct. Geol.* **15**, 1149–1156 (1992).
23. Jóhannesson, H. & Saemundsson, K. *Geological Map of Iceland 1:500,000* (Iceland Mus. of natural Hist. & Iceland geod. Surv., Reykjavik, 1989).

ACKNOWLEDGEMENTS. We thank A. Gudmundsson for his help and advice in the field, and P. Cowie for discussions and comments on the manuscript. We also thank G. Yielding for his comments.

Carbon isotopic evidence for the emergence of C4 plants in the Neogene from Pakistan and Kenya

Michèle E. Morgan*, John D. Kingston† & Bruno D. Marino‡

* Department of Anthropology and ‡ Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

† Department of Anthropology, Yale University, New Haven, Connecticut 06520, USA

DETAILED palaeorecords of terrestrial environments are critical for interpretations of the late Miocene mammalian evolution. Here we assess the development of modern tropical grassland (C4) biomes, an important ecosystem for modern faunas. We present enamel apatite ¹³C/¹²C isotope ratios of fauna from two sequences, the Siwaliks from the Potwar Plateau in northern Pakistan, and the Tugen Hills Succession in Kenya (Fig. 1). Our data do not support previous suggestions of a restricted global date for C4 emergence associated with decreased atmospheric CO₂ around 7 Myr ago¹. Instead, C4 grasses are first recorded as a dietary component in Pakistan at 9.4 Myr and increase as a foraging resource over the next few million years. In Kenya, C4 grasses are present by 15.3 Myr but are not the primary dietary resource of herbivores until 7 Myr. A wide range of C3 and C4 dietary signals characterizes the latest Miocene/Pliocene of Kenya.

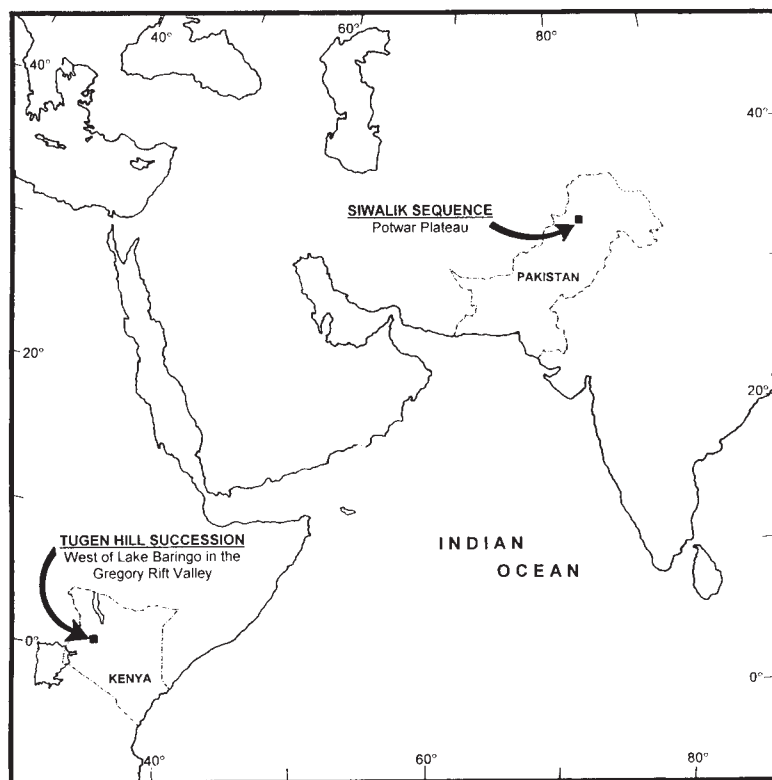
Stable carbon isotope ratios of collagen are valuable dietary indicators in numerous modern and archaeological studies^{2–5}. The application of this technique to the inorganic carbonate fraction of bone and enamel apatite yields consistent results in modern specimens^{6,7} and has been applied to fossil teeth^{1,6,8,9}. This approach is based on the differential selection of atmospheric ¹³CO₂ by the three photosynthetic pathways associated with C3, C4 and CAM (crassulacean acid metabolism) plants (Fig. 2). Mammalian bone and enamel apatite enriches dietary plant $\delta^{13}\text{C}$ by about 12‰ (Fig. 2)^{6,7,10,11}.

Carbonate carbon, liberated as CO₂ for isotopic analysis, primarily substitutes in enamel apatite for phosphate and, much less frequently for hydroxide ions, in three thermally stable posi-

Received 25 June; accepted 4 November 1993.

1. Scholz, C. H. & Cowie, P. A. *Nature* **346**, 837–839 (1990).
2. Walsh, J. J. & Watterson, J. *J. struct. Geol.* **10**, 239–247 (1988).

FIG. 1 Map including East Africa and Southwest Asia showing latitude and longitude of study sites.



tions within the crystalline structure^{12,13}. A fourth labile carbonate position has been identified but is very rare in mature enamel¹³. The validity of results obtained for enamel in this study are contingent on preservation of the *in vivo* signature recorded in the enamel apatite structure.

We have investigated the possibility of diagenetic alteration by measuring the carbon yields both of modern and of fossil specimens subjected to the same pretreatment procedure (Table 1). These data indicate that the weight per cent (wt %) carbon of our fossil enamel specimens is similar to that of modern enamel and is neither contaminated with large amounts of exogenous carbon compounds, nor depleted in carbon. Weight per cent carbon may be a useful criterion for selecting fossil enamel specimens. We are currently investigating post-mortem changes in crystal structure and exchange in an effort to assess more precisely the stability of carbonate ions.

Our fossil data from Pakistan and Kenya generally support previous analyses that in the latest Miocene and Pliocene (<7 Myr) C4 grasses were extensive in areas of both southern Asia and eastern Africa^{14,16} (Fig. 3). The Siwalik enamel data record a shift between 9.5 and 5 Myr from C3- to C4-based

vegetation whereas the Tugen Hills data indicate increased heterogeneity of C3/C4 habitats through the late Miocene and Pliocene. But our data from both regions suggest that the emergence of C4 grasses was more complicated than proposed by others^{1,14,15}.

In the Siwaliks there is dietary evidence for C4 grasses at 9.4 Myr, significantly earlier than the onset of the palaeosol carbonate transition at 7.4 Myr. Independent evidence for palaeodiet from dental microwear analyses of several Siwalik artiodactyl and perissodactyl species included in this isotopic study indicate that grasses were an important dietary component of many herbivores by 9.5 Myr, and were present locally by at least 14 Myr¹⁷. Therefore, in contrast to previous models¹⁴, C3 grasses were an integral part of the Siwalik vegetation before evidence for the appearance of C4 grasses at 9.4 Myr. For example, $\delta^{13}\text{C}$ values of 16 equid specimens span 10‰ (−13.5 to −3.7‰) over a 3.1 Myr period. Identified by dental microwear as intermediate browser-grazers, the varied isotopic values of equids suggest a diet incorporating both C3 and C4 grasses, probably dependent on availability. Although C4 grasses become more prevalent in the latest Miocene and early Pliocene,

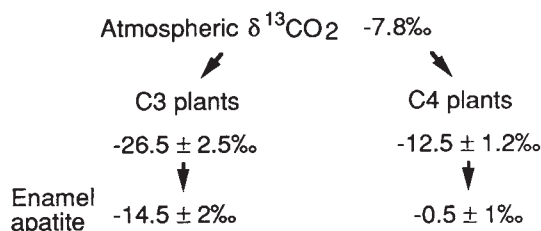
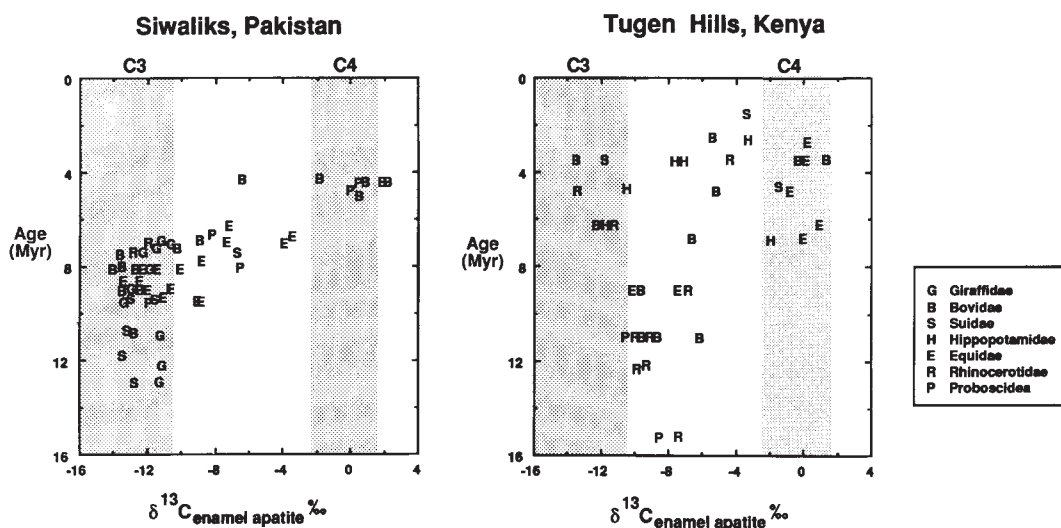


FIG. 2 Schematic summary of carbon pathway. $\delta^{13}\text{C}$ is defined as the ratio of the $^{13}\text{C}/^{12}\text{C}$ of the sample relative to the ratio of the conventional Pee Dee Belemnite (PDB) standard expressed in parts per thousand ($\delta^{13}\text{C} = \{[(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{PDB}}] - 1\} \times 10^3$ ‰). Atmospheric $\delta^{13}\text{C}$ currently averages −7.8‰ (ref. 27); pre-industrial values were −6.5‰ (ref. 28). Extant C3 plants, which select strongly against $^{13}\text{CO}_2$, include most trees, shrubs and herbs, as well as grasses preferring cool, wet growing seasons^{29,32}. Extant C4 plants are almost exclusively grasses favouring hot, dry growing seasons and discriminate less against $^{13}\text{CO}_2$ than C3 plants. CAM plants use both the C3 and the C4 pathways and may have tissue carbon isotope ratios between −9 and −34‰. CAM plants primarily include xeric-adapted flora and are assumed to have comprised an insignificant fraction of the plant biomass during the period we studied although virtually nothing is known about the evolution of the CAM photosynthetic pathway. Modern enamel apatite values are from ref. 6.

FIG. 3 $\delta^{13}\text{C}$ values of Siwalik ($n=58$) and Tugen hills herbivores ($n=40$). Stippled areas indicate likely ranges of pure C3 and pure C4 diets. For Siwalik specimens giraffids, G, include two species, *Giraffokeryx punjabiensis* (older than 9 Myr) and *Bramatherium megacephalum* (younger specimens). Equids, E, probably represent a single species before 8.2 Myr, and 2 species in younger levels. Suids, S, include two species, *Listriodon pentapotamiae* (before 10 Myr), and *Hippopotamodon sivalense* (younger specimens). Bovids, B, include several specimens identified to species. P, gomphotheres, an extinct family of proboscideans, and R, rhinocerotids from the family Rhinocerotidae. Five specimens are not noted because of extensive overlap (B, 9.1 Myr, $\delta^{13}\text{C} = -12.3\text{‰}$; B, 9.1 Myr, $\delta^{13}\text{C} = -12.6\text{‰}$; E, 9.4 Myr, $\delta^{13}\text{C} = -8.9\text{‰}$; E, 9.0 Myr, $\delta^{13}\text{C} = -16.6\text{‰}$; P, 4.4 Myr, $\delta^{13}\text{C} = 0.8\text{‰}$). Note the consistent C3 signal of all giraffids; both species are browsers based on analysis of morphology and dental microwear. Tugen Hills specimens are primarily identified to family with several specimens identified to species. Proboscideans, P, may include specimens from several families.



METHODS. All samples were prepared by treating cleaned and powdered enamel with 2% sodium hypochlorite and 1 M acetic acid, successively for 24 h. Pure phosphoric acid was reacted with samples at 90 °C for 24 h. Resultant CO_2 was collected by cryogenic separation and analysed on a VG prism mass spectrometer. Precision was $\pm 0.1\text{‰}$ for $\delta^{13}\text{C}$ ratios of 5 replicate pairs of fossil enamel. Two internal standards run during the period of analysis yielded standard deviations of $\pm 0.08\text{‰}$ ($n=17$) and $\pm 0.05\text{‰}$ ($n=33$). On the basis of these data, the overall external precision was within 0.1‰.

a high-crowned bovid from 4.3 Myr ($\delta^{13}\text{C}$ of -6.5‰) indicates that C3 grasses periodically comprised a significant portion of the plant biomass.

All giraffid $\delta^{13}\text{C}$ values are consistent with a C3 browsing diet as supported by dental microwear analysis¹⁷. Despite the wide range of $\delta^{13}\text{C}$ values of other specimens between 6 and 8.2 Myr, the tight clustering of the five giraffid specimens of *Bramatherium megacephalum*, collected from different fossil localities, argues against diagenetic alteration of carbonate.

In the Tugen Hills succession, C4 plants were present by 15.3 Myr as recorded by proboscidean and rhinocerotid tooth enamel. These data are the earliest faunal indicators of the presence of C4 plants and are supported by mixed C3–C4 signals from palaeosol carbonates and pedogenic organic matter from the same sequence and time interval¹⁶. Fossil grasses preserved at Fort Ternan, Kenya, dated to 14 Myr, may have included C4 species¹⁸. The Tugen Hills data first record exclusive C4 diets at 6 Myr. Five equids ($0.1 \pm 0.6\text{‰}$) analysed from sediments younger than 7 Myr cluster around pure C4 values. Mixed C3–C4 vegetation as reflected by enamel apatite $\delta^{13}\text{C}$ values between 7 and 2 Myr appears to have supported a wide range of possible dietary regimes.

Marked faunal change occurred in both sequences during the latest Miocene (<8.5 Myr). In the Siwaliks many woodland-dependent fauna, including hominoids, became locally extinct and predominantly grazing high-crowned bovids and equids came to dominate faunal assemblages^{19–21}. In the Tugen Hills,

species associated with modern East African fauna replaced archaic middle Miocene taxa^{22,23}. The character of the latest Miocene faunas in both sequences is consistent with our isotopic indicators of more open and seasonal environments.

The isotopic evidence for C4 grasses in these two geographical areas is relevant to questions concerning their spatial and temporal origin. It has been suggested that a dramatic decrease in atmospheric CO_2 levels between 11 and 8 Myr²⁴ may have favoured more efficient photosynthetic use of CO_2 and, therefore, C4 plants over C3 plants²⁵. Although the increased importance of C4 plants in the latest Miocene supports this idea¹, the dietary presence of C4 grasses by 9.4 Myr in Pakistan, and 15.3 Myr in Kenya, suggests that C4 plants were stable members of plant communities for millions of years before the advent of widespread C4-based savannah grasslands. In addition, isotopic and microwear data from the same species and localities in Pakistan indicate that C3 grasses were more abundant components of the vegetation during the middle and late Miocene than anticipated from earlier interpretations of climatic models. We suggest that C4 grasses evolved gradually in favourable ecosystems in response to a number of environmental and physiological factors affecting photosynthetic production²⁶, rather than primarily as a response to late Miocene changing atmospheric conditions. □

Received 1 June; accepted 19 October 1993.

1. Cerling, T. E., Wang, Y. & Quade, J. *Nature* **361**, 344–345 (1993).
2. Ambrose, S. H. & DeNiro, M. J. *Oecologia* **69**, 395–406 (1986).
3. DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495–506 (1978).
4. Schoeninger, M. J. & Moore, K. J. *World Prehistory* **6**, 247–296 (1992).
5. Van der Merwe, N. J. & Vogel, J. C. *Nature* **276**, 815–816 (1978).
6. Lee-Thorp, J. & van der Merwe, N. J. *S. Afr. J. Sci.* **83**, 712–715 (1987).
7. Lee-Thorp, J. A., Sealy, J. C. & van der Merwe, N. J. *J. Archaeol. Sci.* **16**, 585–599 (1989).
8. Koch, P. L., Zachos, J. C. & Gingerich, P. D. *Nature* **358**, 319–322 (1992).
9. Quade, J. et al. *Chem. Geol.* **94**, 183–192 (1992).
10. Kreuger, H. W. & Sullivan, C. H. *Stable Isotopes in Nutrition*. Am. Chem. Soc. Symp. Ser. 258 (eds Turnland, J. R. & Johnson, P. E.) 205–220 (American Chemical Society, Washington DC, 1984).
11. Sullivan, C. H. & Kreuger, H. W. *Nature* **292**, 333–335 (1981).
12. Beshah, K., Rey, C., Glimcher, M. J., Shimizu, M. & Griffin, R. G. *J. Solid State Chem.* **84**, 71–81 (1990).
13. Rey, C., Renugopalakrishnan, V., Shimizu, M., Collins, B. & Glimcher, M. J. *Calcif. Tissue Int.* **49**, 259–268 (1991).
14. Quade, J., Cerling, T. E. & Bowman, J. R. *Nature* **342**, 163–166 (1989).

TABLE 1 Inorganic carbon yields (wt %) of modern and fossil herbivores

Modern ($N=2$)	1.22 ± 0.18
Siwalik ($N=49$)	1.29 ± 0.20
Tugen Hills ($N=44$)	1.16 ± 0.26

Data are consistent with literature values^{13,33}. Yields are shown as mean ± 1 s.d.

15. Cerling, T. *Palaeogeogr. Palaeoclim. Palaeoecol.* **97**, 241–247 (1992).
16. Kingston, J. D., Hill, A. & Marino, B. *Am. J. Phys. Anthropol.* **14A**, 100 (1992).
17. Morgan, M. E. & Solounias, N. *J. Vert. Paleo.* **10**, 36A (1990).
18. Dugas, D. P. & Retallack, G. J. *J. Paleont.* **67**, 113–128 (1993).
19. Barry, J. C., Lindsay, E. H. & Jacobs, L. L. *Palaeogeogr. Palaeoclim. Palaeoecol.* **37**, 95–130 (1982).
20. Barry, J. C. et al. *Paleobiology* **17**, 231–245 (1991).
21. Flynn, L. J. & Jacobs, L. L. *Palaeogeogr. Palaeoclim. Palaeoecol.* **37**, 129–138 (1982).
22. Hill, A. *Anthropologie* **89**, 275–279 (1985).
23. Hill, A. *J. Hum. Evol.* **16**, 583–598 (1987).
24. Freeman, K. H. & Hayes, J. M. *Global Biogeochem. Cycles* **6**, 185–198 (1992).
25. Ehrlinger, J. R., Sage, R. F., Flanagan, L. B. & Percy, R. W. *Trends Ecol. Evol.* **6**, 95–99 (1991).
26. Peisker, M. *Pl. Cell Envir.* **9**, 627–635 (1986).
27. Keeling, C. D. et al. *Am. Geophys. Monogr.* No 55 (ed. Peterson, D. H.) 165–236 (American Geophysical Union, Washington DC, 1989).
28. Marino, B. D. & McElroy, M. B. *Nature* **349**, 127–131 (1991).
29. Deines, P. in *Handbook of Environmental Isotope Geochemistry: The Terrestrial Environment* Vol. 1 (eds Fritz, P. & Fontes, J. C.) 329–406 (Elsevier, New York, 1980).
30. Farquhar, G. D., O'Leary, M. H. & Berry, J. C. *Aust. J. Pl. Physiol.* **9**, 121–137 (1982).
31. O'Leary, M. H. *Phytochemistry* **20**, 553–567 (1981).
32. Vogel, J. C. *Fractionation of the Carbon Isotopes During Photosynthesis* (Springer, Berlin, 1980).
33. Brudevdof, F. & Soremark, R. in *Structural and Chemical Organization of Teeth* Vol. II (ed. Miles, A. E. W.) 247–277 (Academic, New York, 1967).

ACKNOWLEDGEMENTS. We thank D. Pilbeam, J. C. Barry and A. Hill for their support of this project and for comments on drafts of this letter. Funding was provided by the Leakey Foundation, the National Science Foundation, the Smithsonian Institution Foreign Currency Program, and the Wenner Gren Foundation for Anthropological Research.

Preference for symmetric males by female zebra finches

John P. Swaddle & Innes C. Cuthill

School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

FLUCTUATING asymmetries are the small random deviations from symmetry that occur in the development of bilaterally symmetrical traits¹. Such asymmetries are believed to be a direct indicator of phenotypic, perhaps genotypic, quality². Relative levels of asymmetry are much greater in secondary sexual characters than in normal morphological traits³. It has been proposed that females use the amount of asymmetry in secondary sexual ornamentation as a cue in mate choice³, because the highest quality individuals have the most symmetrical, as well as elaborate, ornaments^{2–10}. Using a choice chamber similar to that in ref. 11, we show here that female zebra finches choose symmetrically leg-banded males over asymmetrically banded ones. This demonstrates unequivocally that females use symmetry as a criterion in partner preference, although whether the symmetry preference is specific to secondary sexual characters is unknown.

Fluctuating asymmetries (FAs) are thought to arise from the inability of individuals to buffer themselves against environmental or genetic stresses during development^{2,12}. There are intraspecific differences in the extent to which individuals resist these developmental accidents, and these individual differences are heritable^{10,12}. 'Good genes' models of sexual selection predict that mate choice should be based on traits that honestly reflect heritable fitness gains¹³; it has been suggested that FAs provide just such information^{2,5}. It is therefore significant that bilaterally symmetrical secondary sexual ornaments show greater FAs than ordinary morphological characters, and are sufficiently large to be detected by eye^{2,7}. Furthermore, asymmetry is negatively correlated with ornament size^{4,7}, implying that high quality individuals not only invest more in ornament growth but also resist the developmental stresses that cause asymmetry. But only one study has investigated whether females actually use symmetry as a cue in mate choice. By experimentally manipulating tail length of male swallows, the order of pairing was found to be consistent with mate choice based on both ornament size and symmetry³. But as tail asymmetry impairs flight performance¹⁴, choice may be based on this rather than tail symmetry itself. Here we directly test for an effect of ornament symmetry on partner preference, using an established experimental model in which the signal used in mate choice can be manipulated without confounding effects on flight performance.

Through experimental manipulations, zebra finches (*Taeniopygia guttata*) were found to show sexual preferences for members of the opposite sex that were wearing specific coloured leg bands¹¹. Preferences shown in the choice chamber are highly related to actual mate choice and pairing in free-flight aviaries^{15–22}. Birds banded with 'attractive' colours tended to have higher reproductive success¹⁶, a higher proportion of same-sex offspring^{18,22}, lower parental investment per offspring and lower mortality rates¹⁵. All the leg-band combinations were symmetrical for left and right legs. By manipulating the symmetry of the leg-band combinations, while keeping the amount of each colour constant between treatments, we investigated whether female zebra finches prefer symmetrically banded males. Although males were shown to have specific colour preferences for banded females^{11,15–22}, for simplicity our experiment is limited to female choice.

Ten male and ten female wild-type zebra finches of roughly the same age were obtained from two different breeders, to ensure that stimulus and test birds were not familiar with each other. Every bird was ringed with a numbered orange leg band (colour bands were supplied by A. C. Hughes, Middlesex, UK, as in ref. 11). The test apparatus was similar, but not identical, to that used in ref. 11. Our apparatus included a sheet of one-way glass at the end of every stimulus compartment (Fig. 1). This

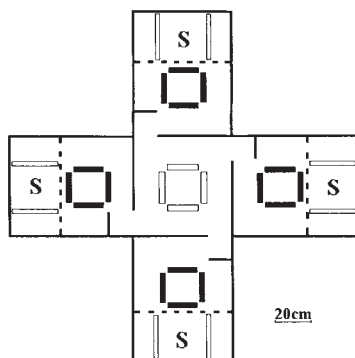
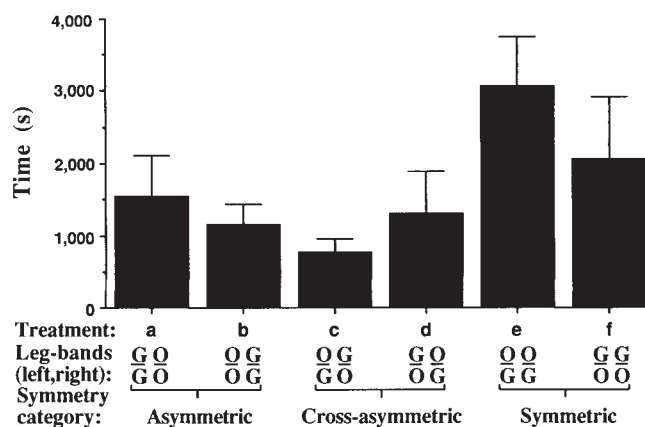


FIG. 1 The choice chamber, plan view. Hollow bars represent fixed perches; filled bars represent microswitch perches; dashed lines represent one-way glass; S, stimulus cage. Background room illumination was provided by two 80 W fluorescent tubes 1.5 m above the apparatus. In addition, an 8 W fluorescent tube was placed at the back of each stimulus cage to provide an increasing illumination gradient from test to stimulus cages and to ensure correct functioning of the one-way glass. Females were pre-exposed to the apparatus for a total of 6 h each before the experimental trials began. The 10 experimental females were those out of a pool of 14 birds that moved most actively within the apparatus during this pre-experimental phase. The birds, when not in the experimental apparatus, were housed in two same-sex cages with abundant food and water at a constant temperature of 18 °C on a 14:10 h light:dark photoperiod. The cages were visually, but not acoustically, isolated. Following the protocol in ref. 11, they also received passive exposure to the leg bands by placing a cage of non-experimental males (from the same suppliers as the experimental males) that were wearing the experimental bands facing the cage of females. This reduces any neophobic reaction to novel leg coloration, and perhaps allows initial acquisition of band preferences¹¹. Colours were rotated between these non-experimental males each day.

FIG. 2 The mean time ($\pm$ s.e.) spent facing males of each colour band treatment. Note that, owing to presence of a central 'neutral' compartment (Fig. 1), these times need not sum to the trial length; on average, females spent 37% of the 5 h trial in front of males. Transformation of the data was not required. Analysis was by nested mixed-model (repeated measures) ANOVA, with 'female' as a random effect, 'symmetry category' as a three-level fixed effect (symmetric, asymmetric and cross-symmetric), and ring position (a versus b, c versus d, e versus f) as a fixed effect nested within symmetry category. The treatments are defined in the text and figure legend, with ring colours O, orange and G, green. Females spent a significantly greater amount of time in front of symmetrically banded males than asymmetrically or cross-asymmetrically banded males ($F_{2,18} = 6.09$; $P = 0.010$; group differences tested by orthogonal contrasts (ref. 29); symmetric versus asymmetric and cross-asymmetric, $t = 3.39$, $P = 0.003$; asymmetric versus cross-asymmetric, $t = 0.81$, $P = 0.501$). Within symmetry categories there was no preference for particular ring positions ($F_{3,27} = 0.81$, $P = 0.501$).



ensured that the stimulus males could not see the test female, and so negated any effects that female display characteristics might have on male behaviour.

At the beginning of each experimental trial, four male (stimulus) birds were placed in the peripheral cages of the apparatus, one in each cage (Fig. 1). A female (test) bird was placed in the centre of the apparatus and allowed to fly freely between these stimulus compartments. All stimulus birds wore two green and two orange leg bands so that treatments differed only in symmetry, not in total colour. Four bands of two colours, with two on each leg, can be arranged in six combinations: two asymmetric (left and right leg different colours); two cross-asymmetric (each leg has both colours, but relative positions differ); and two symmetric treatments (each leg has both colours in the same relative positions). Within each of these experimental categories were the two possible mirror-image alternatives (see legend to Fig. 2), allowing us to deconfound effects of ring position from symmetry.

The experiment consisted of thirty trials, during which all test females were used three times and all stimulus males were used twelve times. The experimental design was balanced so that all males appeared in each treatment group an equal number of times, all 15 of the possible 4-way permutations of the 6 treatments (a to f) were presented, and all treatment groups appeared in every stimulus cage position an equal number of times. Thus, although only 10 stimulus males wore the rings, any preference for ring combinations cannot be confounded with preference for a particular male, a particular stimulus cage, or a cage position. Females were randomly allocated to trials. Following a 2-hour acclimatization period during which all females visited all males at least once, each trial lasted for 5 hours. The time spent in front of each male was determined by means of microswitch

perches placed in every stimulus compartment, and linked to a computer.

Females spent a significantly greater amount of time in stimulus compartments where there was a symmetric, as opposed to asymmetric or cross-asymmetric, banded male (Fig. 2). The females displayed no ring-position preference within symmetry categories (Fig. 2). These results indicate that female zebra finches can detect and respond to colour pattern symmetry. As the symmetric manipulation was preferred over the asymmetric treatment, it is bilateral symmetry (left leg same as right) that is important rather than within-leg vertical symmetry (upper ring same as lower). Furthermore, as symmetric is preferred over cross-asymmetric, it is not just the average colour of the two legs that must be similar, but the specific pattern arrangement.

These findings also indicate that symmetry preference can be generalized, as manipulations were not altering the symmetry of an existing secondary sexual trait. This can be interpreted in similar ways to the demonstration of ring-colour preference itself^{11,15-22}. Either the rings mimic a specific morphological feature whose colour symmetry is assessed in natural mate choice, or females have a general preference for symmetry that extends to arbitrary characters. In the latter case, the bias may be a by-product of a perceptual system designed for symmetry detection in other contexts (compare with pitch preference in frogs²³, tail-length preference in platyfish²⁴, or emergent properties in general²⁵). Alternatively, symmetry preference may have evolved specifically because symmetry reflects individual condition and is detectable in bilateral secondary sexual traits^{2,3,5}. Whether symmetry preference is widespread, and/or specific to signalling traits is unknown, but it is clear that symmetry is correlated with various measures of fitness^{2,10,26,27}, so there must be co-evolutionary consequences for signal design²⁸. □

Received 7 June; accepted 18 October 1993.

- Ludwig, W. *Das Rechts-Links Problem im Tierreich und beim Menschen* (Springer, Berlin, 1932).
- Møller, A. P. & Pomiankowski, A. *Genetica* (in the press).
- Møller, A. P. *Nature* **357**, 238-240 (1992).
- Manning, J. T. & Hartley, M. A. *Anim. Behav.* **42**, 1020-1021 (1991).
- Møller, A. P. *Anim. Behav.* **40**, 1185-1187 (1990).
- Møller, A. P. *Proc. R. Soc. B* **248**, 199-206 (1992).
- Møller, A. P. & Höglund, J. *Proc. R. Soc. B* **245**, 1-5 (1991).
- Thornhill, R. *Behav. Ecol.* **3**, 277-283 (1992).
- Thornhill, R. *Behav. Ecol. Sociobiol.* **30**, 357-363 (1992).
- Thornhill, R. & Sauer, P. *Anim. Behav.* **43**, 255-264 (1992).
- Burley, N., Krantzberg, G. & Radman, P. *Anim. Behav.* **30**, 444-455 (1982).
- Van Valen, L. *Evolution* **16**, 125-142 (1962).
- Bradbury, J. W. & Andersson, M. B. *Sexual Selection: Testing the Alternatives* (Wiley, Chichester, 1987).
- Møller, A. P. *Proc. R. Soc. B* **243**, 59-62 (1991).
- Burley, N. *Auk* **102**, 647-651 (1985).

- Burley, N. *Am. Nat.* **127**, 415-455 (1985).
- Burley, N. *Anim. Behav.* **34**, 1732-1741 (1986).
- Burley, N. *Evolution* **40**, 1191-1206 (1986).
- Burley, N. *Ethology* **76**, 133-151 (1987).
- Burley, N. *Am. Nat.* **132**, 611-628 (1988).
- Burley, N. *Anim. Behav.* **36**, 1235-1237 (1988).
- Burley, N. *Science* **211**, 721-722 (1981).
- Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. *Nature* **343**, 66-67 (1990).
- Basolo, A. L. *Science* **250**, 808-810 (1990).
- Enquist, M. & Arak, A. *Nature* **361**, 446-448 (1993).
- Russell, E. S. A. N. Y. *Acad. Sci.* **57**, 597-605 (1954).
- Beardmore, J. A. *Heredity* **14**, 411-422 (1960).
- Endler, J. A. & McLellan, T. A. *Rev. Ecol. Syst.* **19**, 395-421 (1988).
- SPSS Inc. *SPSSX User's Guide* 3rd edn (SPSS Inc., Chicago, 1988).

ACKNOWLEDGEMENTS. We thank M. Witter, J. Hutchinson and N. Burley for discussion and comment, A. Pomiankowski and A. Møller for access to unpublished manuscripts, and the SERC for financial support of J.P.S.

Restoration of function by replacement of spinal cord segments in the rat

Yasushi Iwashita^{*†}, Saburo Kawaguchi^{*} & Miyahiko Murata^{*}

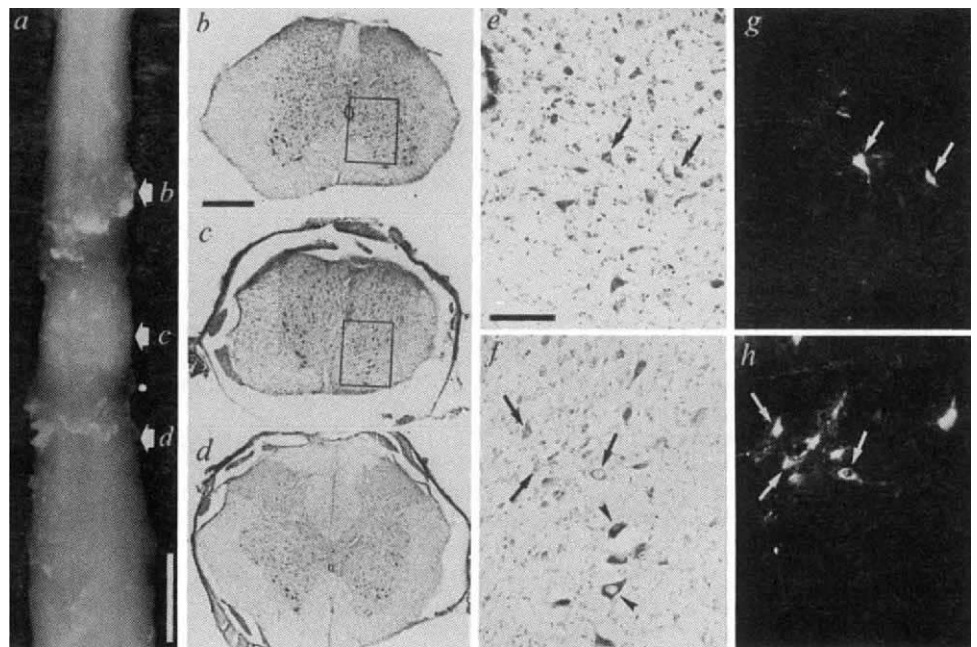
^{*} Department of Integrative Brain Science and [†] Orthopaedic Surgery Faculty of Medicine, Kyoto University, Kyoto 606, Japan

RECONSTRUCTION of a severed mammalian spinal cord with restoration of function has so far not been achieved¹⁻¹², although structural and functional restitution after spinal transection has been successful in some lower vertebrates¹²⁻¹⁴. In quail-chick and chick-chick chimaeras, spinal cord segments were found to be functional after replacement by isotopic and isochronic grafting of the neural tube^{15,16}. Here we achieve such a replacement in neonatal rats under less restricted topological and temporal conditions than were necessary for the avian chimaeras. The replaced segments united with the host spinal cord and promoted robust growth and regrowth of axons across the graft, enabling neural connections to be reconstructed that were hardly distinguishable from normal. The animals with replaced segments could walk, run and climb with almost normal hind-forelimb coordination. This functional restoration in these animals appeared to be permanent, raising the possibility of therapeutic application in humans.

The lower thoracic segments of the spinal cord were resected in neonatal rats and embryonic spinal cord segments were grafted into their place in the normal orientation. Control animals were grafted in the same way but in the inverted dorsoventral or rostrocaudal orientation, or grafted instead with the sciatic nerve, or left ungrafted. The morphology and electrophysiology of the grafts and behaviour of the animals were studied in order to investigate the neural connections across the graft.

Forty-one out of the 60 rats that had undergone surgery survived until histological examination. These included 22 of the 32 rats that had been grafted in the normal orientation, 8 of the 12 grafted in the inverted orientation, 6 of the 8 grafted with the sciatic nerve, and 5 of the 8 ungrafted controls. In 14 of the 22 rats grafted in the normal orientation, the graft united with the recipient spinal cord at both the rostral and caudal stumps and formed a seamless spinal cord (Fig. 1a), whereas in the rest the graft was completely absorbed and extinguished. In six of the former 14 successful cases, the graft had the dorsal and ventral roots and showed the organotypic cytoarchitecture, albeit rather distorted (Fig. 1c). In the remaining eight animals, the graft was mainly composed of white matter (Fig. 2a). In all of these animals no obvious boundaries, such as glial scars, were observed between the graft and the host spinal cord (Fig. 2a, b, g). Many neurons in the upper brain structures and the cervical and thoracic segments including the graft were retrogradely labelled with Fast Blue injected into the lumbar enlargement (Figs 1g, h and 2c-f). The corticospinal projection was anterogradely labelled up to the lumbar enlargement with

FIG. 1 External appearance and cytoarchitecture of the replaced spinal cord segments. *a*, Dorsal view of the spinal cord containing the grafted segments (middle portion) which are thick at the centre and narrower at the host-graft interface. The graft had the dorsal (asterisk) and ventral roots. *b-d*, Photomicrographs of transverse sections at the levels indicated by arrows in (*a*). Compare the distorted but organotypic cytoarchitecture of the graft (*c*) with normal cytoarchitecture of the thoracic (*b*) and lumbar segment (*d*) of the host. Note the disappearance of the central canal in the graft (*c*). *e-h*, The framed areas in (*b*) and (*c*) are shown in (*e*) and (*f*) under bright-field illumination, and in (*g*) and (*h*) under fluorescence illumination, respectively. Photomicrographs (*g*) and (*h*) show neurons that were retrogradely labelled with Fast Blue injected into the lumbar enlargement (arrows). Note the presence of unlabelled neurons in the graft (*f*) with features characteristic of motor neurons (arrowheads). Scale bars: 2 mm in *a*; 500 μ m in *b*, applicable to *b-d*; 100 μ m in *e*, applicable to *e-h*.

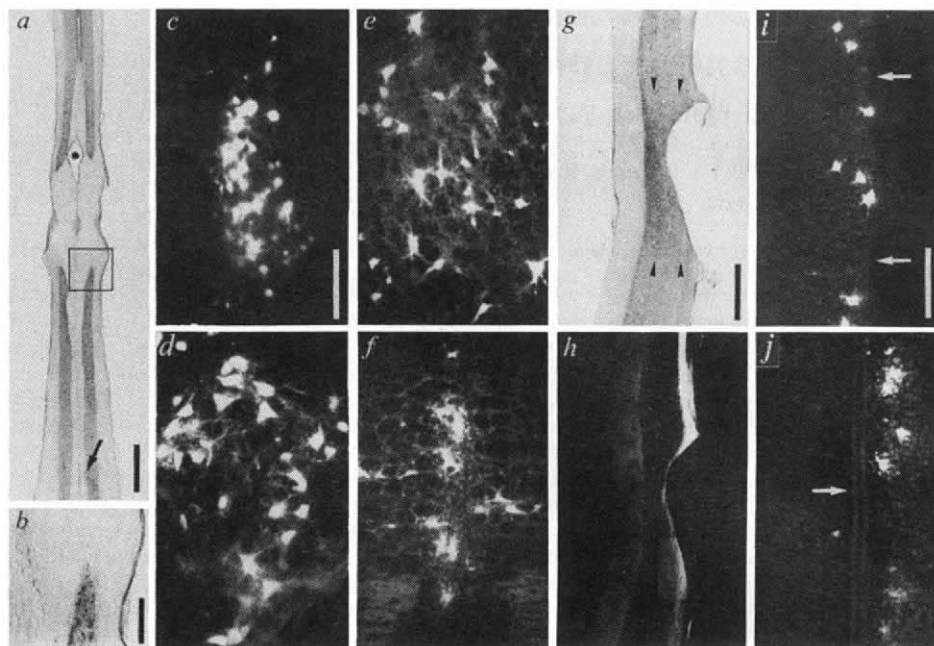


2 mm in *a*; 500 μ m in *b*, applicable to *b-d*; 100 μ m in *e*, applicable to *e-h*. Data were taken from the animal identified by P1-E16-73, in which the age of the host at the time of grafting was postnatal day 1 (P1), the age of the donor was embryonic day 16 (E16), and the time of death was 73 days after grafting. When the days for identification were the same in several animals, they were distinguished by a suffix number, as P1-E16-73-1, P1-E16-73-2 for example.

METHODS. In neonatal rats (1 or 2 days old, Wistar strain), under ether anaesthesia, the lower thoracic segments (T_{10} – T_{11}) of the spinal cord were resected by complete transection followed by aspiration. The transection was made with a razor blade at a distance of 1.5–2.0 mm that corresponds to ~1.5–2 segments. The completeness of resection was confirmed by direct observation of the vacant vertebral canal, in which there was no trace of any filaments. A section of the spinal cord

containing segments T_{10} and T_{11} , equal in length to the resected portion, was dissected from embryos (E14 to E17) taken from the dated pregnant rat under deep pentobarbitone anaesthesia (40 mg kg⁻¹, injected i.p.). The dissected segments were transferred to the vacant place in the vertebral canal of a neonate with apposition of the stumps in the normal orientation. Control animals were grafted with similar segments in the inverted ventrodorsal or rostrocaudal orientation, or grafted with a section of the sciatic nerve of appropriate length taken from an adult rat, or ungrafted (resection only). In these animals, after survival times of 1–4 months, and in normal controls, fluorescent dye (Fast Blue or Diamidino Yellow) was injected into the lumbar enlargement or thalamus (VPL) or cerebellum, and wheat-germ agglutinin-conjugated horseradish peroxidase (WGA-HRP) into the sensorimotor cortex, then the brain and spinal cord were processed histologically²⁴.

FIG. 2 Photomicrographs showing neural connection across the graft. *a*, Horizontal section of the spinal cord containing the grafted segments which were mainly composed of white matter, taken from rat P2-E16-83. Asterisk indicates a cavity at the host-graft interface. Cytoarchitecture of the graft was quite unlike that in Fig. 1 but locomotor function was similar. Arrow in the lumbar enlargement indicates the injection site of Fast Blue; rostral is at the top. *b*, Higher magnification of the framed area in *a* showing no glial scars at the host-graft interface. *c-f*, Fluorescence photomicrographs of the retrogradely labelled neurons in the red nucleus (*c*), vestibular nucleus (*d*), bulbar reticular formation (*e*), and raphe nucleus (*f*), taken from the same rat as in *a*. The presence of these labelled neurons is indicative of regeneration of the transected fibres because most of the brainstem-spinal projections have already reached the lumbar enlargement at birth^{8,25}. *g*, A sagittal section of the spinal cord containing the grafted segments, taken from rat P2-E14-36. The host-graft interface (arrowheads) showed no obvious boundaries such as glial scars. *h*, The same section as in *g* under dark-field illumination showing the pyramidal tract labelled anterogradely with WGA-HRP injected into the sensorimotor cortex. Labelled fibres descended in the dorsal funiculus, traversed the dorsal portion of the graft, and reached the lumbar enlargement. Because the transection was done at a lower thoracic level before the arrival of the corticospinal axons^{7,8,26,27}, these labelled fibres were considered to be formed by developmental growth of uncut fibres. Rostral is at the top. *i*, Horizontal section of the lumbar enlargement under fluorescence illumination showing spinocerebellar neurons²⁸ retrogradely labelled with Diamidino Yellow injected into the cerebellum; taken from rat P2-E16-83-1. *j*, Horizontal section of the lumbar enlargement under fluorescence illumination showing spinothalamic neurons²⁹ retrogradely labelled with Fast



Blue injected into the thalamic VPL nucleus. They were distributed close to the central canal (arrow); taken from rat P2-E16-83-2. Scale bars: 2 mm in *a*, 600 μ m in *b*, 200 μ m in *c* (applicable to *c-f*), 1 mm in *g* (applicable to *g* and *h*), and 100 μ m in *i* (applicable to *i* and *j*). The number of the retrogradely labelled neurons in representative sections from the animal with successful replacement compared with the normal control (grafted/normal animals) was as follows: 45/97 in the red nucleus, 34/69 in the vestibular nucleus, 37/55 in the bulbar reticular formation, and 10/26 in the raphe nucleus. It was estimated that nearly half of those neurons in the normal control send axons to the lumbar enlargement across the graft. The labelled corticospinal projection fibres were traceable to the caudal end of the tissue block that was 7.5 mm caudal to the presumed caudal host-graft interface.

WGA-HRP injected into the sensorimotor cortex (Fig. 2*h*). The spinothalamic and spinocerebellar projection neurons in the lumbar enlargement were retrogradely labelled with fluorescent dyes injected into the respective targets (Fig. 2*i,j*). The neural connections across the graft (Fig. 2) were much the same as normal and were functionally active (Fig. 3). In the animals grafted with the sciatic nerve, injection of Fast-Blue into the lumbar enlargement labelled several neurons in the thoracic segments just rostral to the graft and a negligibly small number of neurons in the bulbar region, but none in the more rostral structures. A similar injection in the ungrafted animals did not label any neurons rostral to the resection.

The animals with successful replacement exhibited good motor performance with hind-forelimb coordination (Fig. 4*a,c*). The limbs in these animals smoothly struck the ground alternately in the order the left hindlimb, left forelimb, right hindlimb, right forelimb, as in normal animals¹⁷. When the animal was dropped in mid-air with limbs pointed upward, it was able to land in the upright position on all four extremities. This indicates that a chain of righting reflexes starting from the rotation of the head acted finally upon the hindlimbs. All animals of the control groups and the unsuccessful experimental group, by contrast, showed functional deficits of various grades: from flaccid paralysis of the hindlimbs with loss of bladder and bowel control (paralytic case) to abnormal walking (paretic case). In the paretic cases, the hindlimbs stepped alternately or, more often, synchronously (Fig. 4*d*). In both types of stepping, the pelvic girdle behaved as an independent unit without coordina-

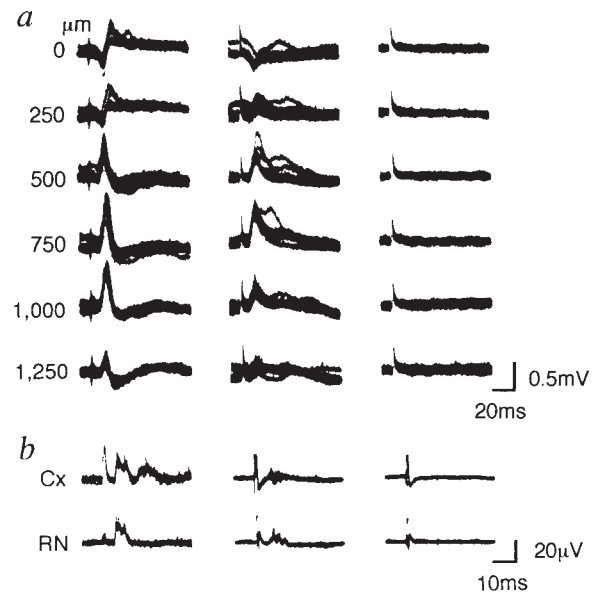
tion of the shoulder girdle. This independence of the pelvic girdle is illustrated in Fig. 4*b*.

In adult animals grafted with a peripheral nerve section into a transected spinal cord, sparse short neural connections were formed across the graft^{3,4} but were not functional³, which is consistent with our findings for control animals with a sciatic nerve graft. In neonatal and adult rats, growth and regrowth of axons were promoted by grafting the embryonic spinal cord into the lesioned spinal cord^{6,11}. The amelioration of functional deficits in neonates after a similar operation has been at least partly ascribed to such promotion¹⁰. Functional restitution after partial transection¹⁰, however, is questionable, because a lesion extending for even as much as 90% of the transverse plane does not produce any paraplegic syndrome in the rat¹¹, presumably owing to extensive reorganization of the spared fibre connections. It must be noted that functional deficits were remarkably mild when the lesion was made in neonates; even after the complete resection of lower thoracic segments, the animals were able to walk. In such cases, however, hind-forelimb coordination was absent (Fig. 4*b*). The nearly normal interlimb coordination in the animals with replaced spinal cord segments after complete resection of lower thoracic segments in the present experiments (Fig. 4*a*), therefore, affords unambiguous evidence that the functional restoration was caused by neural connections reconstructed across the graft.

Growth or regrowth of axons in the central nervous system (CNS) is likely to be affected by their path- and target-finding, which require positional cues¹⁸⁻²⁰. Therefore, the positional

FIG. 3 Synaptic connectivity of the ascending and descending tracts across the graft. *a*, Depth profiles of somatosensory evoked potentials (SEP) in the cerebral cortex on stimulation of the sciatic nerve in a control (left panel) and in a grafted rat (centre). After transection of the spinal cord at a level rostral to the graft, the potentials in the latter animal were abolished completely (right). Numbers to the left of specimen records indicate depths from the cortical surface. *b*, Compound action potentials of the sciatic nerve on stimulation of the cerebral cortex (Cx) and the red nucleus (RN) in the control (left) and in the grafted rat (centre). The potentials in the latter animal were abolished completely after spinal cord transection (right). Disappearance of the potentials in the cerebral cortex and the sciatic nerve by spinal cord transection indicates that both potentials were evidently mediated by synaptic connectivity across the graft. All potentials were recorded on the side contralateral to stimulation. Each specimen record consists of seven superposed traces. Voltage calibrations were upward negative. The control animal was a 72-day-old intact rat; the grafted rat was P2-E15-78. The potentials in the grafted rat were very similar to those in the control animal, although the magnitude was smaller in the grafted rat than in the control. The SEP on stimulation of the sciatic nerve³⁰, a surface positive-depth negative wave with reversal of potential at a depth ~250 μm from the cortical surface, was restricted in the hindlimb area in both the grafted and control rats. There was no difference in the latency of response between the grafted and control rats, whereas the time-to-peak of the depth negative wave of the SEP was slightly longer in the grafted (16 ms) than in the control (14 ms).

METHODS. Two rats with replaced spinal cord segments and two intact rats of matched age for control were used. Under sodium pentobarbitone anaesthesia (40 mg kg^{-1} ; i.p.), the trachea and the femoral vein were cannulated. Sodium pentobarbitone was intermittently supplemented intravenously to maintain anaesthesia, the level of which was monitored by electrocorticogram. Respiration was mostly natural, but was artificial for a short time as a result of immobilization by pancuronium dibromide (0.08 mg kg^{-1} ; i.v.) to confirm that responses were free

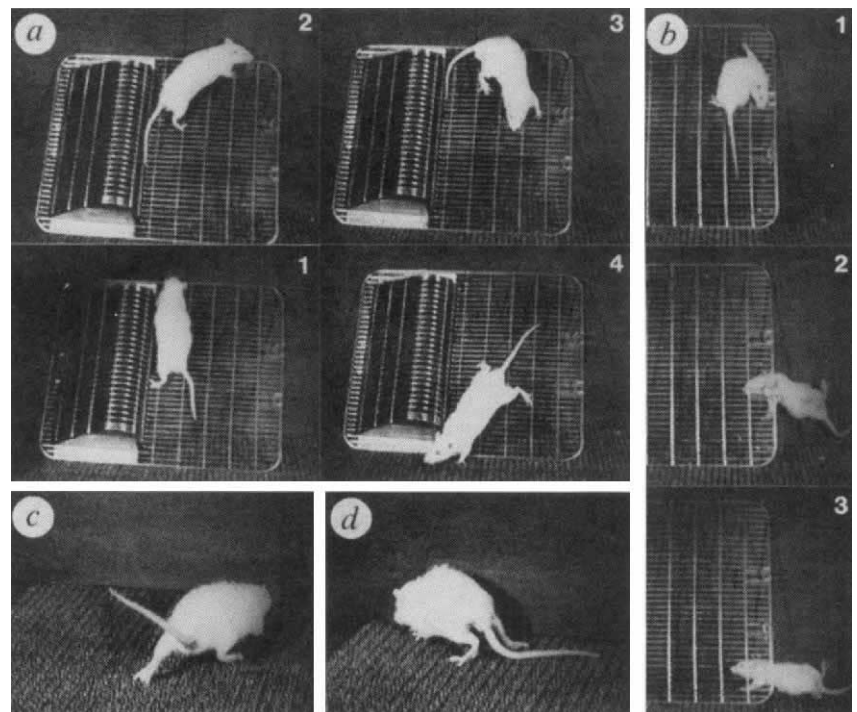


from movement-associated artefact. The sciatic nerve was cut and mounted on a bipolar electrode, which was used for both stimulation and recording. After recording the SEP, a pair of bipolar electrodes were placed in the cortex and the red nucleus for stimulation, and compound action potentials were recorded from the sciatic nerve. Then the spinal cord was transected to confirm that the responses were mediated by tracts across the graft. Operative procedures and methods of stimulation and recording were similar to those described²⁴.

information in the graft appears to be important for growth and regrowth of CNS axons. This may explain the success of the normally oriented grafts compared with the poor performance

of the inversely oriented ones and sciatic nerve grafts. It is consistent with the finding in the avian chimaeras that replacement of spinal cord segments was only successful when isotopic^{15,16}.

FIG. 4 Videoframes showing locomotor function in the animals with replaced spinal cord segments and in controls. Climbing was tested on the lid of a rat cage standing against a wall and spontaneous walking and running were observed on a floor mat. *a*, Sequential videoframes in the course of climbing up (panels 1 and 2) and down (panels 3 and 4). Taken from the rat used for Fig. 2a-f. The limbs held the wire grid of the lid alternately, with nearly normal hind-forelimb coordination. As the animal turned to the right (from panel 2 to panel 3), the pelvic girdle smoothly followed the rotation of the shoulder girdle. *b*, Sequential videoframes in the course of falling down from the lid. Taken from the ungrafted control P1-resection-83-1. The hindlimbs were able to hold the wire grid, but as the animal turned to the right, they did not follow the rotation of the shoulder girdle, resulting in an abnormal flexion of the body axis (panel 1). Then the four limbs could not hold the body weight and the animal fell in the air (panel 2). During falling, a chain of righting reflexes acted upon the upper body but not on the lower body. Therefore the upper body landed in the normal upright position but the lower body remained supine (panel 3). *c*, A frame taken during walking (same rat as in *a*). The right forelimb and the left hindlimb are extended and the right hindlimb are flexed. There was no swaying of the pelvic girdle, but there was a frequent abnormal myoclonic movement of the tail during walking. *d*, Frame taken during walking (ungrafted control rat P1-resection-83-2). External rotation of the hindlimbs was a typical posture of the rat whose neural connections between the upper and lower body were disconnected. Swaying of the pelvic girdle during walking was remarkable.



Among the control groups, the proportion of the paretic case was higher in the animals grafted with the sciatic nerve (5/6) than in the animals without graft (2/5) and those grafted with spinal cord segments in the inverted orientation (2/8).

In this regard, it is worth noting that the spatial organization of neuronal development is segmentally controlled by homeotic genes which play a part in specifying the positional identity along the rostro-caudal and ventro-dorsal axes^{16,21,22}. In contrast to the widely held view that the mammalian CNS lacks regenerative capacity, it has an incredible potential for regrowth and self-organization^{23,24}. Reconstruction of the severed human spinal cord may not be impossible. □

Received 17 May; accepted 12 November 1993.

1. Ramón y Cajal, S. *Degeneration and Regeneration of the Nervous System* (Hafner, New York, 1959).
2. Björklund, A. *Trends Neurosci.* **14**, 319–322 (1991).
3. Richardson, P. M., McGuinness, U. M. & Aguayo, A. J. *Nature* **284**, 264–265 (1980).
4. Kao, C. C., Chang, L. W. & Bloodworth, J. M. B. *Jr Exp Neurol.* **54**, 591–615 (1977).
5. Schnell, L. & Schwab, M. E. *Nature* **343**, 269–272 (1990).
6. Reier, P. J., Bregman, B. S. & Wujek, J. R. *J. comp. Neurol.* **247**, 275–296 (1986).
7. Bregman, B. S., Kunkel-Bagden, E., McAtee, M. & O'Neill, A. J. *comp. Neurol.* **282**, 355–370 (1989).
8. Bregman, B. S. & Bernstein-Goral, H. *Expl Neurol.* **112**, 49–63 (1991).
9. Sieradzian, K. & Vrbová, G. *Neuroscience* **31**, 115–130 (1989).
10. Kunkel-Bagden, E. & Bregman, B. S. *Expl Brain Res.* **81**, 25–34 (1990).

11. Das, G. D. in *Neural Transplantation and Regeneration* (eds Das, G. D. & Wallace, R. B.) 1–61 (Springer, New York, 1986).
12. Guth, L., Reier, P. J., Barrett, C. P. & Donati, E. J. *Trends Neurosci.* **6**, 20–24 (1983).
13. Selzer, M. E. *J. Physiol., Lond.* **277**, 395–408 (1978).
14. Singer, M., Nordlander, R. H. & Egar, M. J. *comp. Neurol.* **185**, 1–22 (1979).
15. Kinutani, M. & Le Douarin, N. M. *Dev Biol.* **111**, 243–255 (1985).
16. Le Douarin, N. M. *Trends Neurosci.* **16**, 64–72 (1993).
17. Grillner, S. *Physiol. Rev.* **55**, 247–304 (1975).
18. Harris, W. A. *Nature* **339**, 218–221 (1989).
19. Tessier-Lavigne, M., Placzek, M., Lumsden, A. G. S., Dodd, J. & Jessell, T. M. *Nature* **336**, 775–778 (1988).
20. Bolz, J., Götz, M., Hübener, M. & Novak, N. *Trends Neurosci.* **16**, 310–316 (1993).
21. Kessel, M. & Gruss, P. *Science* **249**, 374–379 (1990).
22. Itasaki, N. & Nakamura, H. *Neuron* **8**, 787–798 (1992).
23. Kalil, K. & Reh, T. *Science* **205**, 1158–1161 (1979).
24. Kawaguchi, S., Miyata, H. & Kato, N. *J. comp. Neurol.* **245**, 258–273 (1986).
25. Leong, S. K., Shieh, J. Y. & Wong, W. C. *J. comp. Neurol.* **228**, 18–23 (1984).
26. Schreyer, D. J. & Jones, E. G. *Neuroscience* **7**, 1837–1853 (1982).
27. Bernstein, D. R. & Stelzner, D. J. *J. comp. Neurol.* **221**, 382–400 (1983).
28. Matsushita, M. & Hosoya, Y. *Brain Res.* **173**, 185–200 (1979).
29. Giesler, G. J., Jr, Spiel, H. R. & Willis, W. D. *J. comp. Neurol.* **195**, 243–252 (1981).
30. Amassian, V. E., Waller, H. J. & Macy, J. Jr *Ann. N.Y. Acad. Sci.* **112**, 5–32 (1964).

ACKNOWLEDGEMENTS. We thank K. Sasaki for encouragement, advice and criticism, N. Kato for critically reading the manuscript, and Y. Kurimoto for technical advice. Supported by the Ministry of Education, Science and Culture of Japan on Grant-in-Aid for Scientific Research and Grant-in-Aid for Project Research.

Neurotrophin-3 enhances sprouting of corticospinal tract during development and after adult spinal cord lesion

Lisa Schnell, Regula Schnelder, Roland Kolbeck*, Yves-Alain Barde* & Martin E. Schwab

Brain Research Institute, University of Zurich, August Forel-Strasse 1, CH-8029 Zurich, Switzerland

* Department of Neurobiochemistry, Max-Planck-Institute for Psychiatry, 82152 Martinsried by Munich, Germany

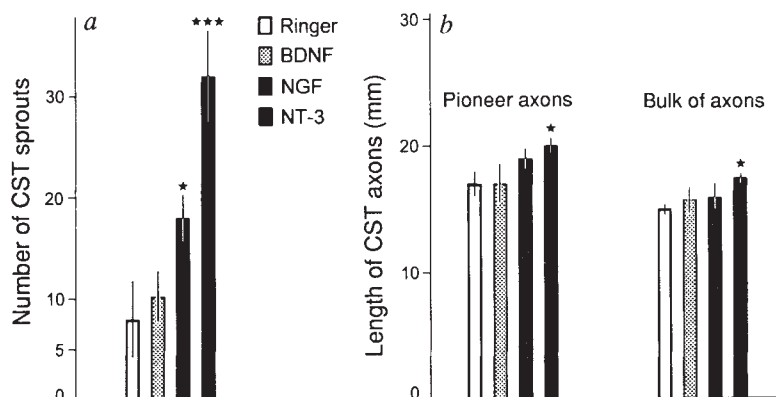
THE number of neurotrophic factors found in the central nervous system is rapidly growing, but their functions *in vivo* are largely unknown. In the peripheral nervous system they promote the survival of developing and lesioned neurons and enhance nerve fibre growth and regeneration^{1–6}. Here we study the effects of nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF) and neurotrophin-3 (NT-3) on the largest tract system leading from the brain to the spinal cord, the corticospinal tract (CST)*. The developing CST grows down the spinal cord during the first

postnatal days and innervates its targets after a waiting period by collateral sprouting^{8–10}. We find that NT-3 injected locally specifically enhances this sprouting, whereas BDNF has no effect. In adult rats, injection of NT-3 (but not BDNF) into the lesioned spinal cord increases the regenerative sprouting of the transected CST. The distance of growth of the sprouts is very restricted, but application of an antibody that neutralizes myelin-associated neurite growth inhibitory proteins¹¹ results in long-distance regeneration of CST fibres.

We investigated the effects of neurotrophic factors on the developing rat CST by injecting 0.8–1 µg NT-3, BDNF or NGF or 1 µl Ringer solution into the lower thoracic spinal cord of postnatal-day-2 (P2) rats. At P2, CST axons have entered the upper thoracic spinal cord; they will reach the injection site at about P3 (refs 8–10). Rats were killed at P5, a time when most of the CST axons are still growing and collateral sprouting has barely started^{8–10}. CST fibres were labelled one day before by injection of the tracer wheat-germ agglutinin-horseradish peroxidase (WGA-HRP) into the sensory motor cortex. The labelling intensity of the CST was similar in all the groups. The extension of the main CST axons was measured directly, taking the vertebral level C1 as a landmark. A small enhancement of growth was visible with NT-3 but not with BDNF (Fig. 1b). Sprouting of side branches was quantified in the thoracic spinal cord (level T6) by counting all the labelled CST branches intersecting a dorso-ventral line (that is, a virtual transvers plane) on complete

FIG. 1 *a*, Collateral sprouting, and *b*, growth of the main axons of the developing rat CST at postnatal day 5. NT-3 and, to a lesser degree, NGF enhance the sprouting of CST collaterals (*a*). Only a minor effect is present on the growth rate of the primary axons (*b*).

METHODS. 1 µg NT-3 (murine recombinant; produced and purified according to ref. 27; 1 µg µl⁻¹), BDNF (murine recombinant; produced and tested for biological activity as described in ref. 28), NGF (mouse 2.5S NGF) or vehicle (Ringer solution) alone was injected without laminectomy using a Hamilton syringe fitted with a glass capillary into the lower thoracic spinal cord of P2 Lewis rats (3–6 rats per experimental condition). 0.5 µl WGA-HRP (Sigma; 5% in water) were injected into the sensory motor cortex unilaterally at P4. Rats were fixed by perfusion with buffered 1.25% glutaraldehyde/1% formaldehyde 24 h later. 50-micron-thick sagittal frozen sections were reacted for HRP using tetramethylbenzidine as substrate. Sprouts were counted on number-coded and complete series of sections under polarized light in darkfield at a defined level (T6). Numbers in *a* represent means ± s.e.m. of the intersections of sprouts (tracks of grains) with a dorso-ventral line at T6 in each animal. The length of the



primary CST axons was measured taking the vertebral level C1 as a landmark. Level of significance (Student's *t*-test): **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (see Fig. 3).

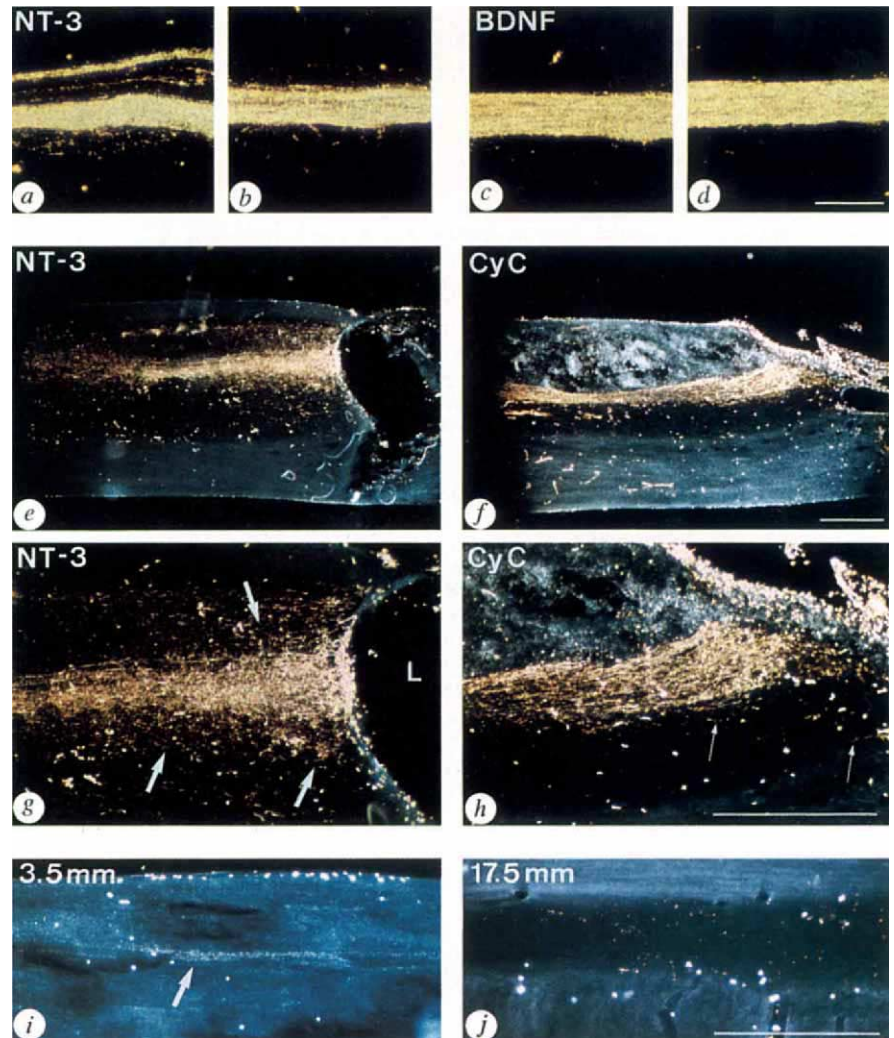


FIG. 2 *a–d*, CST of 5-day-old rats (thoracic level T6; darkfield micrographs). NT-3 induced premature sprouting (*a, b*), in contrast to BDNF (*c, d*). *e–h*, Regenerative sprouting of the transected adult CST rostral to the lesion. *e* and *g*, NT-3 injected spinal cord at low (*e*) and higher magnification (*g*); dense sprouting occurs ventrally and dorsally of the CST (arrows). CST ends at a large cavern at the lesion site (*L*); *f* and *h*, cytochrome *c* (CyC)-injected spinal cord; CST ends at lesion cavern but shows only few sprouts (thin arrow); *i*, bundle of regenerated CST fibres elongating at 3.5 mm caudal to the lesion in the dorsal funiculus of a IN-1/NT-3 treated animal; *j*, single fibres in grey matter of the sacral spinal cord at 17.5 mm caudal to the lesion from the same animal. Each scale bar represents 100 μ in *a–d*, and 500 μ in *e–j*.

series of parasagittal sections of one side of the spinal cord. As shown in Figs 1*a* and Fig. 2*a, b*, the density of CST branches was greatly increased in the NT-3 injected rats; BDNF-treated rats, in contrast, showed the low density of branches typical for that developmental stage and the same as in the Ringer-injected rats (Figs 1*a* and 2*c, d*). Injections of NGF produced an intermediate enhancement of sprouting (Fig. 1*a*).

In response to spinal cord lesions in the adult, CST fibres spontaneously sprout to a low degree, but show no elongation^{12–14}. Sprouting and elongation were studied in response to neurotrophin injections in young adult (4–7 weeks old) rats. The dorsal halves of the spinal cord were transected bilaterally at the lower thoracic level. NGF, BDNF or NT-3 (300–500 μ g in 0.3–0.5 μ l) were injected immediately rostral to the lesion into the spinal cord. Controls received equivalent amounts of cytochrome *c* or vehicle (Ringer solution). In half of the experiments, hybridoma cells producing the monoclonal antibody IN-1 (ref. 11) or a control antibody (anti-HRP antibody; no inhibition of HRP enzymatic activity¹³) were injected into the left parietal cortex and lateral ventricle. The antibodies secreted by the locally forming tumours reached the spinal cord via the cerebrospinal fluid¹³. Fourteen to 17 days later, the CST axons were traced by anterograde transport of WGA–HRP from the right sensory-motor cortex.

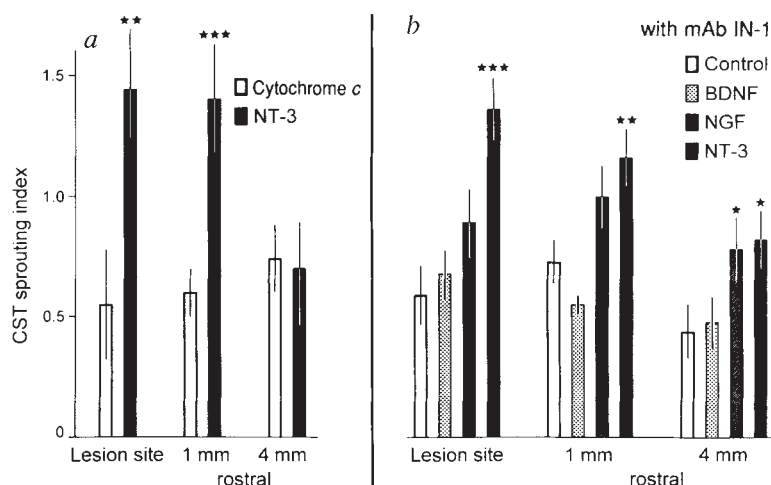
Regenerative sprouting was quantified on number-coded complete parasagittal section series of each spinal cord by counting all the labelled CST branches intersecting vertical lines at the lesion site, 1 mm rostral and 4 mm rostral to the lesion. To

account for the variability in the anterograde tracing and CST labelling in these adult rats, the numbers of sprout intersections were divided by the total number of labelled axons within the compact CST at the corresponding levels in each individual rat (branching index). The index value for the normal collateral branching present in unlesioned animals was subtracted in order to obtain a specific sprouting index (given by the number of regenerative sprouts per labelled CST axon)¹⁴. Spontaneous sprouting of the lesioned CST fibres occurred at all three levels (Fig. 3*a*). A single injection of NT-3 greatly increased this sprouting at the lesion site and at 1 mm (Figs 2*e, g* and 3*a*). Control injections with cytochrome *c* (Figs 2*f, h* and 3*a*) or Ringer solution alone (Fig. 3*b*) were indistinguishable. No significant effect of antibody IN-1 could be detected except at 4 mm, where the effect of NT-3 was enhanced by IN-1 (Fig. 3*a, b*). Comparisons of three neurotrophins showed significant differences, very similar to their effects on the developing CST; BDNF had no effect on sprouting when compared to controls, in contrast to the highly significant enhancement of sprouting seen with NT-3 (Fig. 3*b*). NGF enhanced CST sprouting but not significantly except at 4 mm (Fig. 3*b*).

The number of labelled axons in the CST varied from animal to animal (mainly for technical reasons related to the tracing procedure), but the values overlapped for all the groups, and the mean numbers for each group were very similar. This result shows that the neuronal metabolism, as shown by anterograde transport of WGA–HRP, was not measurably affected by the trophic factor treatment.

FIG. 3 Regenerative sprouting of the transected, adult CST. *a*, Sprouting is enhanced by NT-3. Sprouting after injection of cytochrome *c* (a protein with physical-chemical properties similar to NT-3, NGF and BDNF) is identical to vehicle alone ('control' in *b*) and to rats with lesion only¹⁴. Sprouting was quantified on 3 levels: at of the lesion site, 1 mm and 4 mm rostral to it. *b*, In contrast to NT-3, the related neurotrophin BDNF has no effect. The smaller effect of NGF is only significantly different from the controls at 4 mm. Factor-induced sprouting is enhanced by the monoclonal antibody (mAb) IN-1 at 4 mm.

METHODS. Lesion: The dorsal halves of the lower thoracic spinal cord were transected bilaterally in 4–7-week-old Lewis rats, thereby completely lesioning the main CST in the dorsal funiculus and its very minor component in the dorso-lateral funiculus. NT-3, BDNF, NGF or cytochrome *c* (Sigma) (0.3–0.5 μ g in 0.3–0.5 μ l of Ringer solution) or vehicle alone were injected rostral to the lesion using a glass capillary. mAb application: hybridoma tumours were implanted in the parietal cortex as described^{13,14}. Survival time after the lesion was 2–3 weeks. The presence of tumours or tumour scars was verified in each animal. Tracing: WGA–HRP was injected (3–4 injections of 0.3–0.5 μ l each) into the sensory motor cortex (on the non-tumour side) 24 hours before killing. Quantification of sprouting: see text and ref. 14. Results shown



are means $\pm$ s.e.m., $N=15-21$ for each group. For significance levels, see legend to Fig. 1.

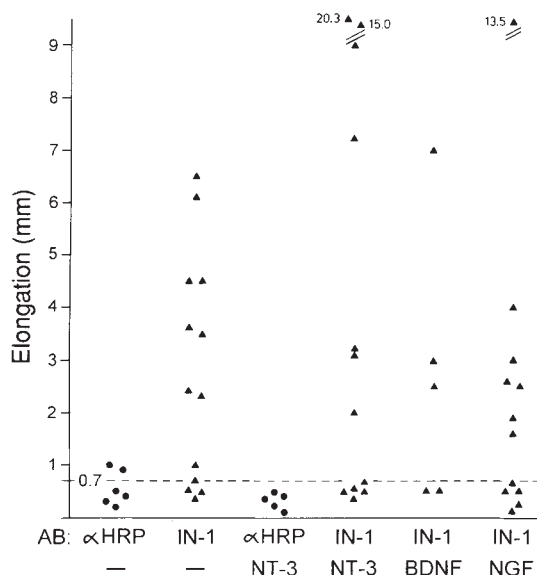
The elongation of the regenerating CST fibres was determined on the same serial parasagittal section series. The distance of the most caudal labelled fibres in each rat was measured in millimetres from the transection site and the results are shown in Fig. 4. Regeneration distances of 0.2–0.7 mm correspond to the mean length of sprouts growing towards and around the lesion site (Fig. 2*e-h*). In the presence of a control antibody (anti-HRP), fibres rarely grew longer than this sprouting distance (Fig. 4). Interestingly, this was also the case in the presence of NT-3, namely under conditions of strongly stimulated sprouting. In contrast, in the presence of antibody IN-1 most rats showed elongation of CST fibres over several millimetres (Figs 2*i, j* and 4). The group receiving the combined treatment of NT-3 and IN-1 contained some animals showing regeneration over very long distances (Figs 2*j* and 4). Although a quantification of the number of regenerating CST axons was not possible in this WGA–HRP-labelled material, mainly because of the fasciculation of the fibres, we estimate that 5–10% of the CST fibres show long-distance growth in the IN-1 rats. Although variable from animal to animal, the distances of regeneration in the NT-3/IN-1 group were the longest we ever observed. In these rats, the regenerating fibres reached the lower lumbar and sacral spinal

cord. Branching of regenerating fibres seemed to occur in the grey matter (Fig. 2*j*), and preliminary results showed functional improvements of CST-mediated sensory motor tasks¹⁵. These results show that stimulation of sprouting by NT-3 is not sufficient to lead to elongation, but regenerating fibres can grow over several millimetres when myelin-associated inhibitory proteins are neutralized by the IN-1 antibody.

Our results show a selective enhancement by NT-3 of branch formation from CST axons during development *in vivo*, as well as in the lesioned adult spinal cord. Of all the neurotrophins, NT-3 shows the highest level of messenger RNA expression in the embryonic and developing spinal cord¹⁶⁻¹⁸. NT-3 mRNA is mainly located in motor neurons¹⁷ and their local density could determine the relative levels of NT-3 in different spinal cord regions. Indeed, CST collateralization is highest in the brachial and lumbar segments¹⁹, which give rise to the innervation of the limbs and where motor neuron numbers are highest²⁰. NT-3, therefore, qualifies as a possible inducer and regulator of CST collateral sprouting, an important step to target-finding and innervation^{8,9,21}. A possible direct effect of NT-3 on the CST fibres is suggested by the fact that the mRNA for *trkC*, a functional receptor for NT-3, is highly expressed in the deeper layers

FIG. 4 Elongation of CST fibres exceeding the sprouting distance of 0.7–1 mm (dashed line) is present only in mAb IN-1-treated rats. The strong stimulation of sprouting by NT-3 is not leading to elongation in the presence of a control mAb (anti-HRP). Symbols show the position of the longest regenerated fibres in each rat, measured in mm from the transection site. IN-1 animals without elongations >0.7 mm may have insufficient levels of mAb, or scars in the lesion area which inhibit fibres from crossing the lesion.

METHODS. The same number-coded sagittal section series were used as for determination of sprouting. Rats with total lesions having no tissue bridges were discarded (39/107 rats), as were rats without hybridoma tumours (15/68 rats). Only fibres that could be followed and reconstructed over long distances were accepted as CST fibres. Regenerated fibres crossed the lesion site on remaining tissue bridges, mostly in regions that are abnormal for CST fibres. The presence of large necrotic areas and caverns lead to a typical complicated, irregular fibre path. In contrast to an earlier study¹³, animals in which CST sprouts did not reach the caudal side of the lesion were also included in our evaluation.



of the developing cortex, where the cells of origin of the CST are found²². Lower levels of trkC mRNA and NT-3 high-affinity binding sites persist in the corresponding layers of the cortex in the adult^{23,24}.

In the lesioned adult spinal cord, NT-3 enhances regenerative sprouting of transected CST fibres, in contrast to BDNF and more than NGF. Interestingly, the stimulation of CST sprouting by NT-3 did not by itself lead to elongation of these fibres over more than 1 mm. It has been claimed that sprouting and elongation of lesioned fibres are differentially regulated processes^{3,25}. Our results strongly support this concept. In the adult CNS, restriction seems to be due to negative effects of the local tissue environment: that of the myelin-associated neurite-growth-inhibitory proteins which can be neutralized by antibody IN-1, and possibly that of other inhibitory components associated with the lesion and scar area²⁶. Growth-enhancing factors such as the neurotrophins and neurite growth inhibitors therefore seem to be important regulators of fibre growth and regeneration in the CNS. □

Received 29 July; accepted 28 October 1993.

1. Barde, Y.-A. *Neuron* **2**, 1525–1534 (1989).
2. Davies, A. M. & Lumsden, M. E. A. *Rev. Neurosci.* **13**, 61–73 (1990).
3. Diamond, J., Holmes, M. & Coughlin, M. J. *Neurosci.* **12**, 1454–1466 (1992).

4. Oppenheim, R. W., Prevette, D., Quin-Wei, Y., Collins, F. & MacDonald, J. *Science* **251**, 1616–1618 (1991).
5. Oppenheim, R. W., Quin-Wei, Y., Prevette, D. & Yan, Q. *Nature* **360**, 755–757 (1992).
6. Sendtner, M., Kreutzberg, G. F. & Thoenen, H. *Nature* **345**, 440–441 (1990).
7. Nudo, R. J. & Masterton, R. B. J. *comp. Neurol.* **277**, 53–79 (1988).
8. Schreyer, D. J. & Jones, E. G. *Neuroscience* **7**, 1837–1853 (1992).
9. O'Leary, D. D. M. & Terashima, T. *Neuron* **1**, 901–910 (1988).
10. Schwab, M. E. & Schnell, L. *J. Neurosci.* **11**, 709–722 (1991).
11. Caroni, P. & Schwab, M. E. *Neuron* **1**, 85–96 (1988).
12. Bregman, B. S., Kunkel-Bagden, E., McAtee, M. & O'Neill, A. J. *comp. Neurol.* **282**, 355–370 (1989).
13. Schnell, L. & Schwab, M. E. *Nature* **343**, 269–272 (1990).
14. Schnell, L. & Schwab, M. E. *Eur. J. Neurosci.* **5**, 1156–1171 (1993).
15. Kunkel-Bagden, E. et al. *Neurosci. Abstr.* **19**, 681 (1993).
16. Maisonnier, P. C. et al. *Neuron* **5**, 501–509 (1990).
17. Ernfrors, P. & Persson, H. *Eur. J. Neurosci.* **3**, 953–961 (1991).
18. Friedman, W. J., Olson, L. & Persson, H. *Eur. J. Neurosci.* **3**, 688–697 (1991).
19. Casale, E. J., Light, A. R. & Rustioni, A. J. *comp. Neurol.* **278**, 275–286 (1988).
20. Romanes, G. J. *Progr. Brain Res.* **11**, 93–116 (1964).
21. Heffner, C. D., Lumsden, A. G. S. & O'Leary, D. D. M. *Science* **247**, 217–220 (1990).
22. Ringstedt, T., Lagercrantz, H. & Persson, H. *Dev. Brain Res.* **72**, 119–131 (1993).
23. Merlio, J.-P., Ernfrors, P., Jaber, M. & Persson, H. *Neuroscience* **51**, 513–532 (1992).
24. Altar, C. A., Criden, M. R., Lindsay, R. M. & DiStefano, P. S. *J. Neurosci.* **13**, 733–743 (1993).
25. Ramon y Cajal, S. *Degeneration and Regeneration of the Nervous System* (Hafner, New York, 1959).
26. Schwab, M. E., Kapfhammer, J. P. & Bandtlow, C. E. A. *Rev. Neurosci.* **16**, 565–595 (1993).
27. Götz, R., Kolbeck, R., Lottspeich, F. & Barde, Y.-A. *Eur. J. Biochem.* **204**, 745–749 (1992).
28. Dechant, G. et al. *Development* **119**, 545–558 (1993).

ACKNOWLEDGEMENTS. We thank R. Lindegger for technical assistance, J. Kapfhammer for critically reading the manuscript and S. Kaufmann for secretarial help. The study was supported by a grant from the International Spinal Research Trust, the American Paralysis Association, the Swiss NSF, the International Research Institute for Paraplegia and Regeneron Pharmaceuticals, Inc.

Apparent motion determined by surface layout not by disparity or three-dimensional distance

Zijiang J. He & Ken Nakayama

Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA

THE most meaningful events ecologically, including the motion of objects, occur in relation to or on surfaces¹. We run along the ground, cars travel on roads, balls roll across lawns, and so on. Even though there are other motions, such as flying of birds, it is likely that motion along surfaces is more frequent and more significant biologically. To examine whether events occurring in relation to surfaces have a preferred status in terms of visual representation, we asked whether the phenomenon of apparent motion would show a preference for motion attached to surfaces. We used a competitive three-dimensional motion paradigm² and found that there is a preference to see motion between tokens placed within the same disparity as opposed to different planes. Supporting our surface-layout hypothesis, the effect of disparity was eliminated either by slanting the tokens so that they were all seen within the same surface plane or by inserting a single slanted background surface upon which the tokens could rest. Additionally, a highly curved stereoscopic surface led to the perception of a more circuitous motion path defined by that surface, instead of the shortest path in three-dimensional space.

Using a stereoscope, observers were presented with a three-dimensional (3-D) apparent motion display that is a modification of one commonly used in 2-D displays². Figure 1a depicts a 2-D motion display that consists of two diagonally placed tokens presented simultaneously, alternately presented with tokens located on the opposite diagonals. Such a competitive motion display will result in a bi-stable motion perception: either vertical or horizontal motion depending on the ratio of inter-token distances in the two directions. For example, when the vertical distance is kept the same, shorter horizontal distances will result in horizontally perceived motion (Fig. 1b), and longer distances will lead to vertically perceived motion (Fig. 1c). This

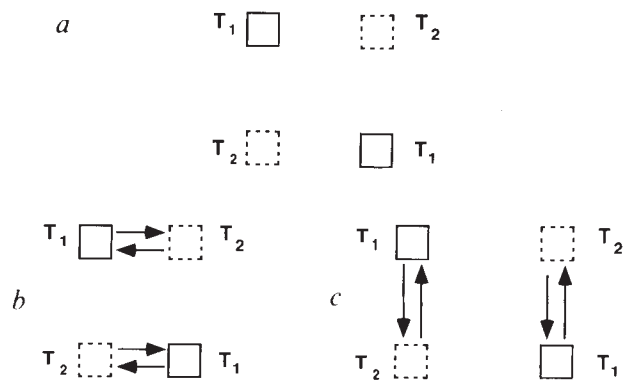
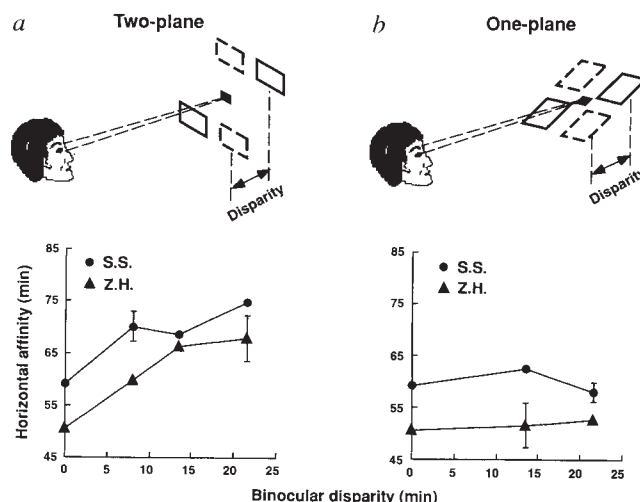


FIG. 1 Bi-stable apparent motion display where stationary targets are presented alternately at T_1 and T_2 (a). The relative proximity between stimuli in successive frames biases the proportion of horizontally versus vertically perceived motion. Thus horizontal motion is more likely to be seen for short horizontal distances (b), whereas vertical motion is more likely to be seen for longer distances (c).

'proximity' tendency³ of the motion token to match its nearest neighbour can be summarized by determining the horizontal distance (horizontal affinity) where there is no bias for either horizontal or vertical motion. The larger the horizontal affinity, the stronger is the motion bias between the horizontal tokens. As it is a competitive apparent motion situation, a larger horizontal affinity also reflects a weaker vertical motion bias between the top and bottom tokens.

In our first experiment, the bottom and top pair of tokens had differing amounts of binocular disparity so that they could appear in the front and back vertical planes, respectively (Fig. 2a; two-plane condition). During the experiments, we systematically increased and decreased the horizontal distance obtaining motion judgments for each presentation. From the percentages of horizontal motion judgments for each horizontal distance, the horizontal affinity at each disparity was obtained by probit analysis⁴. Our result shows that the horizontal affinity increases with increasing disparity (Fig. 2a), indicating that the tendency for seeing motion within the same disparity pair of tokens increases as the disparity between the pairs of tokens increases. This result is essentially identical to those reported by Green and

FIG. 2 Comparison of results obtained from a display where increasing disparity leads to the placement of tokens in either one of two fronto-parallel planes (a) or where the individual tokens are slanted in depth so that they are seen to lie on a single plane, slanting back (b). The results indicate that as binocular disparity is increased, an increased horizontal affinity between the tokens occurs only in the two plane condition. Observers (Z.J.H., and S.S., a naive observer) viewed a computer-generated display at a distance of 100 cm through a pair of haploscopic prisms to obtain stereopsis. For each trial, the two frames of the motion display were presented alternately for 333 milliseconds for a total of six frames (three repetitions of each diagonal pair). Each token (43×21 arc min) was white (70 cd m^{-2}) and viewed against a grey (14 cd m^{-2}) background and a small black fixation square was placed midway in depth between the upper and lower pair of tokens. The observer's task was to maintain fixation and report the apparent motion direction (either horizontal or vertical) after each trial. Each block (for a given binocular disparity) consisted of 96 trials where the horizontal distances were increased and decreased repetitively. Before each session all subjects had over 100 practice trials to reduce hysteresis^{2,13}. Error bars represent the mean standard error obtained for a given curve.



Odom⁵, who suggested an extension of the proximity tendency to 3-D space. Thus greater 3-D distances accompanying increases in binocular disparity weaken the vertical and thus strengthen horizontal motion. Alternatively, one might consider binocular disparity itself to be a primitive feature providing a more distinctive label for determining motion direction.

Our surface-layout hypothesis, however, accounts for the result in terms of surfaces: increasing the disparity also more effectively separates the upper and lower pairs of tokens in terms of distinct non-coplanar surface planes. Horizontal affinity increases because motion within a surface plane is preferred.

To distinguish the effects of disparity or 3-D depth from our surface-layout hypothesis, we made measurements of motion perception under similar conditions of varying disparity but changed the stimulus configurations, so all tokens were perceived to lie in the same surface plane. Instead of maintaining the tokens in a fronto-parallel orientation, the small rectangular tokens were slanted stereoscopically so that all would all lie in a single imaginary plane tilted back (Fig. 2b). Our surface-layout hypothesis predicts less change despite increasing disparity because the top and bottom tokens still lie on a common surface plane. Using the same experimental procedure, we then compared motion perception here with the previous two-plane condition (Fig. 2a). The results were clear. Consistent with the surface-layout hypothesis, there is essentially no increase in the horizontal affinity despite the increasing differentiation of the top and bottom tokens in terms of binocular disparity and/or 3-D distance (Fig. 2b).

In the experiment shown in Fig. 2b, we generated a single slanted surface 'implicitly' by altering the 3-D orientation of the individual tokens so that they could be perceived as coplanar. In our next experiment we sought to generalize our findings by examining the role of a continuously perceived surface and using tokens that did not lie within the surface but which simply appeared to 'rest' on the surface. We used the original set of fronto-parallel tokens in the two-plane condition (Fig. 2a) and placed them in relation to two types of surface backgrounds. First, like the fronto-parallel condition of Fig. 2a, we used a condition where tokens were viewed against a large vertical fronto-parallel surface which had the same disparity as the farthest set of tokens (Fig. 3a). Second, we also used the same exact set of fronto-parallel tokens but had them rest on a receding surface defined stereoscopically as a white background with random dots (Fig. 3b). According to our surface-layout hypothesis, we expected that the increase in horizontal affinity with increasing disparity might be greatly attenuated, particularly in comparison to the case where there was no common plane along which the tokens could be perceived to move. Thus increasing

disparity would increase the horizontal affinity only in the two-plane condition, but barely in the other. The findings shown in Fig. 3 support this view.

In addition, we have discovered a rather different phenomenon that also supports the surface-layout view, showing that the presence of a surface can radically alter the detailed trajectory of apparent motion even though the starting and ending positions of the motion tokens remain the same. We used similar motion stimuli as in the first experiment (one-plane condition), except that the four tokens were drawn to reside within an imaginary curved surface in 3-D (Fig. 4). When our subjects were presented with the tokens alone in the absence of other surfaces, the subjective motion pattern perception in the vertical direction was that of two pairs of moving rectangular tokens, resembling 'paddles',

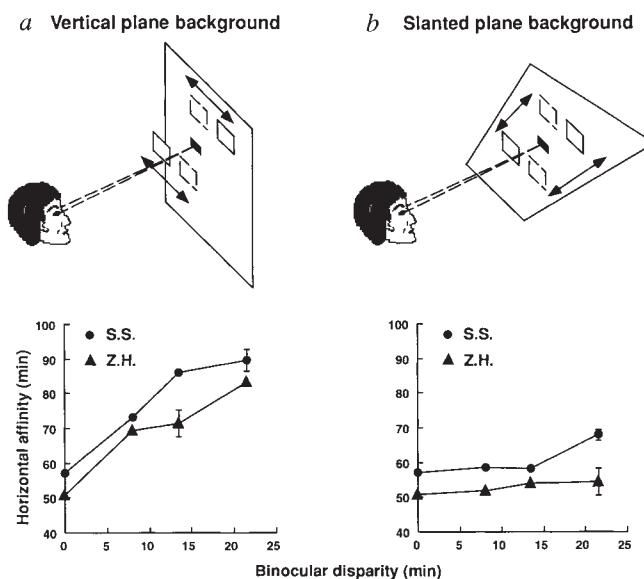
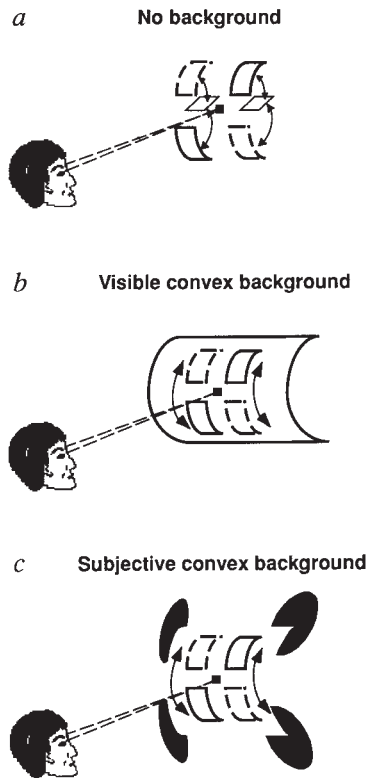


FIG. 3 The effect of visible surface backgrounds on apparent motion correspondence. Tokens here were essentially identical to the ones used in the two-plane condition of the first experiment. They were either placed against a visible vertical background (a), or a slanted background (b). The grey vertical background was rectangular (224×114 arc min), with a mean luminance of 60 cd m^{-2} , and filled with black random dots to enhance depth perception. To generate the slanted background, a similar random-dot background was differentially skewed (in each eye) to form monocular parallelograms such that when viewed stereoscopically, a receding plane was seen.

FIG. 4 Surface guided motion paths: the effect of surface background configurations on the perceived apparent motion trajectories. In each configuration, the curved and slanted tokens had exactly the same spatial configuration, constructed to lie within a semicylindrical convex surface. In *a*, the surface itself was invisible and the motion resembled paddles, flipping vertically. In *b*, this cylindrical surface is visible, defined by a grey background (60 cd m^{-2}) with black random dots having the appropriate disparity. In *c*, the cylindrical surface is subjective, defined by appropriate disparities at the corners, where it is perceived to occlude four disks. In *b* and *c*, the perceived motion path is qualitatively different from *a*, appearing to lie 'within' the cylindrical surface.



flipping across the shortest 3-D path (Fig. 4a). But the presence of a curved surface, defined either by a grey surface with random dots (Fig. 4b) or subjective contours⁶ (Fig. 4c) led to a perceived motion path that was altogether different. Our subjects often saw sliding motion along the more circuitous path defined by the curved surface; that is, it was perceived to move 'within' the surface. We also note that the two different paths are accompanied by very different sets of correspondence rules between the edges of the tokens. With the paddle motion, the nearer and farther edges of one token correspond to the nearer and farther of the other, respectively. In the case of within-surface motion, the near edge of the token corresponds to the farther of the other and vice versa.

These experimental results taken as a whole indicate that the perception of motion is decisively affected by the perception of surfaces. Motion that is seen to be moving in relation to surfaces is favoured even if it means taking a longer path in 3-D space.

Our results argue against a simple 2-D^{7,8} or 3-D^{5,9} proximity rule for apparent motion. Rather than seeing motion affinity decreasing as a simple function of arbitrary 3-D distances, our work indicates the need for a representation of motion tied to surfaces. It raises the specific question as to whether a more satisfactory proximity rule might be obtained by considering distances (that is, geodesics^{10,11}) within a 2-D surface embedded in 3-D space.

More generally, it raises issues regarding the nature of visual spatial representation itself. In sciences related to vision, it is often implicitly assumed that perceptual processing begins with a 2-D representation of an image followed by successive stages to form a full 3-D representation¹², both for purposes of object representation and spatially oriented behaviour. Our results provide support for an alternative view¹, arguing that the representation of surface layout is a more primitive and a more fundamental visual representation, antecedent to any abstract representation of space in terms of euclidian distances. □

Received 23 August; accepted 22 October 1993.

1. Gibson, J. J. *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, 1979).
2. Ramachandran, V. S. & Anstis, S. M. *Nature* **304**, 529–531 (1983).
3. Burt, P. & Sperling, G. *Psychol. Rev.* **88**, 171–195 (1981).
4. Finney, D. J. *Probit Analysis* (Cambridge Univ. Press, UK, 1971).
5. Green, M. & Odom, J. V. *Science* **233**, 1427–1429 (1986).
6. Harris, J. P. & Gregory, R. L. *Perception* **2**, 235–247 (1973).
7. Ullman, S. *The Interpretation of Visual Motion* (MIT Press, Cambridge, MA, 1979).
8. Mutch, K., Smith, I. M. & Yonas, A. *Perception* **12**, 305–312 (1983).
9. Attneave, F. & Block, G. *Percept. Psychophys.* **12**, 301–307 (1973).
10. Foster, D. H. *Bio. Cyber.* **18**, 81–89 (1975).
11. Carlton, E. H. & Shepard, R. N. J. *math. Psychol.* **34**, 127–188 (1990).
12. Marr, D. *Vision* (Freeman, San Francisco, 1982).
13. Shimojo, S. & Nakayama, K. *Perception* **19**, 285–299 (1990).

ACKNOWLEDGEMENTS. This work was supported in part by a grant from Air Force Office of Scientific Research.

Platelet-activating factor as a potential retrograde messenger in CA1 hippocampal long-term potentiation

Kunio Kato*, Gary D. Clark†, Nicolas G. Bazan†‡ & Charles F. Zorumski*§||

Departments of * Psychiatry and § Neurobiology, Washington University School of Medicine, 4940 Children's Place, St Louis, Missouri 63110, USA

† Neuroscience Center and Departments of † Neurology and

‡ Ophthalmology, Louisiana State University School of Medicine, New Orleans, Louisiana 70112, USA

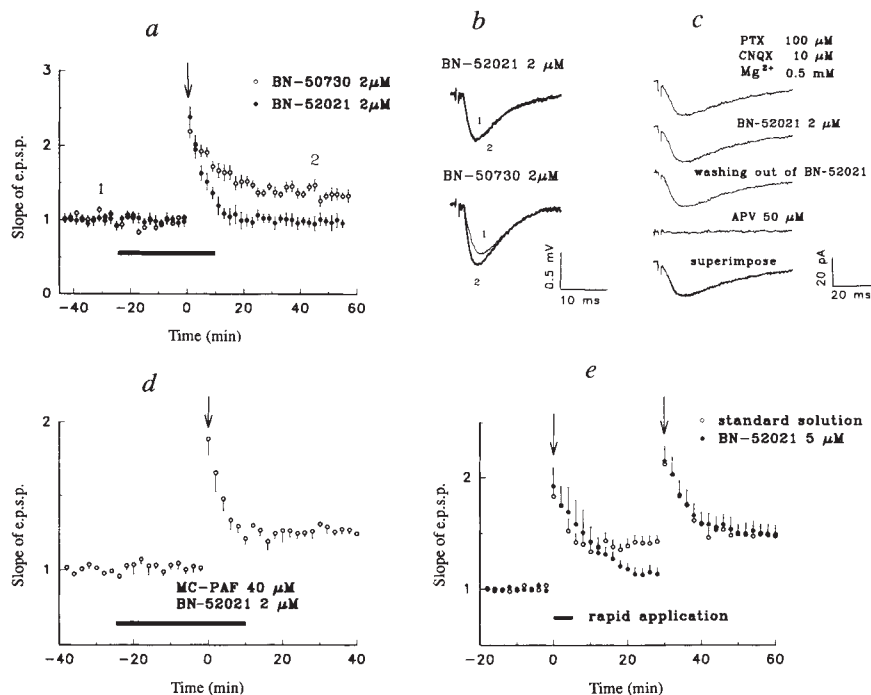
LONG-term potentiation (LTP) refers to a persisting enhancement of neurotransmission that follows high-frequency activation of certain synapses¹. Although both pre- and postsynaptic mechanisms contribute to LTP^{2,3}, it is believed that the enhanced release of neurotransmitter that accompanies this process results from the production of a diffusible messenger in postsynaptic neurons which traverses the synaptic cleft and alters the function of presynaptic terminals^{4,5}. One candidate for such a messenger is arachidonic acid, a metabolite produced by phospholipase A₂ which augments synaptic transmission when coupled with presynaptic stimulation⁵. However, the effects of arachidonic acid require activation of the postsynaptic receptor for *N*-methyl-D-aspartate⁶. Previously we found that platelet-activating factor (1 *O*-alkyl-2-acetyl-*sn*-glycero-3-phosphocholine), another phospholipase A₂-derived messenger⁷, selectively enhances excitatory postsynaptic currents in hippocampal neurons by a presynaptic mechanism⁸. We now present evidence that platelet-activating factor, acting at a receptor localized to synaptic regions, participates in LTP in the CA1 region of rat hippocampal slices and may serve as part of a retrograde signalling cascade.

To determine whether platelet-activating factor (PAF) receptors participate in CA1 LTP, we used bath applications of the PAF antagonists BN-52021 and BN-50730. When administered for 25 min before and for 10 min after tetanic stimulation, 2 μM BN-52021, a synaptosomal PAF antagonist⁹, blocked the enhancement of excitatory postsynaptic potentials (e.p.s.ps) that follows tetanic stimulation of the Schaffer collateral pathway (Fig. 1a, b). Although an initial potentiation was seen in BN-52021-treated slices, responses declined to baseline within 10–20 min after the tetanus. In contrast, BN-50730, a microsomal PAF antagonist¹⁰, failed to inhibit LTP ($34 \pm 9\%$ ($n=5$) and $28 \pm 8\%$ ($n=3$) increase in e.p.s.p. slope at 2 μM and 5 μM

|| To whom correspondence should be addressed.

FIG. 1 A synaptosomal PAF-receptor antagonist blocks CA1 LTP. **a**, BN-52021 (2 μ M, filled circles; $n=10$) blocked the potentiation of field e.p.s.ps, but 2 μ M BN-50730 (open circles; $n=5$) did not. Drugs were applied during the time denoted by the solid bar and the tetanus was supplied at the time shown by the arrow. Results in this and all figures represent mean $\pm$ s.e.m. **b**, The traces depict e.p.s.ps obtained at the times denoted by the numbers in **a**. **c**, BN-52021 does not inhibit the NMDA component of synaptic currents. Each trace represents the average of three currents obtained under the conditions specified. **d**, A high concentration of MC-PAF overcomes the LTP-blocking effect of BN-52021. MC-PAF (40 μ M) and BN-52021 (2 μ M) were applied during the time denoted by the solid bar ($n=4$). **e**, BN-52021 blocks LTP when applied after the tetanus. The filled circles show the slope of e.p.s.ps in slices treated with 5 μ M BN-52021 applied immediately after the tetanus ($n=9$) and the open circles show the responses during a similar application of extracellular solution ($n=7$). The solutions were applied from a cannula located 200 μ m from the recording pipette during the time shown by the solid bar. Tetanic stimulation was supplied at the times shown by the arrows.

METHODS. Hippocampal slices were prepared according to standard procedures. Male albino rats (22–30 days old) were anaesthetized with halothane and decapitated. Hippocampi were rapidly dissected and placed in gassed (95% O₂–5% CO₂) extracellular solution containing in (in mM): 124 NaCl, 5 KCl, 2 CaCl₂, 2 MgSO₄, 1.25 NaH₂PO₄, 22 NaHCO₃ and 10 glucose at 10 °C. Transverse slices (500–600 μ m thick) were cut with a rotary tissue slicer, then maintained in an incubation chamber for at least 2 h at 30 °C. At the time of an experiment, individual slices were transferred to a submersion recording chamber, where they were continuously perfused with extracellular solution (2 ml min⁻¹) at 30 °C. Field recordings were obtained from the apical dendritic region of CA1 using 3–8-M Ω electrodes filled with 2 M NaCl. During an experiment, the Schaffer collateral–commissural fibres were stimulated every 20 s in the stratum radiatum using a concentric bipolar electrode (25 μ m tip, Rhodes Medical Instruments) and 0.3-ms constant-current pulses at an intensity sufficient to evoke a 50% maximal response. LTP was produced by an electrical tetanus administered



for 1 s at 100 Hz using the same stimulus intensity. Data were analysed using an IBM-based system. BN-52021 and BN-50730 were gifts from P. Braquet. Whole-cell recordings were obtained from neurons in the cell body layer of CA1 using the 'blind' patch clamp method³⁰ and an Axopatch-1D amplifier. Patch electrodes were pulled from 1.2-mm outside diameter borosilicate glass and had resistances of <6 M Ω after fire polishing. Pipettes were routinely filled with a solution containing (in mM): 130 caesium methanesulphonate, 10 tetraethylammonium chloride, 5 NaCl, 1 MgCl₂, 0.5 1,2-bis(2-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (BAPTA), 10 *N*-[1-hydroxyethyl]-piperazine-*N'*-[2-ethanesulphonic acid] (HEPES), 2 Mg-ATP and 0.3 Na-GTP, with pH adjusted to 7.25 using CsOH. The NMDA component of synaptic responses was isolated using an extracellular solution containing 0.5 mM Mg²⁺, 100 μ M picrotoxin (PTX) and 10 μ M 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX)³⁰.

respectively). The effects of BN-52021 were not mediated by blockade of *N*-methyl-D-aspartate (NMDA) receptors, as 2 μ M BN-52021 had no effect on the NMDA component of CA1 synaptic currents (Fig. 1c). The LTP-blocking effect of 2 μ M BN-52021 was partially reversed by high concentrations of the PAF analogue methylcarbamil-PAF (MC-PAF) (Fig. 1d). In slices treated with 40 μ M MC-PAF plus 2 μ M BN-52021, there was a $22 \pm 4\%$ increase in e.p.s.p. slope measured 20–30 min after the tetanus ($n=4$). In contrast, control slices exhibited a $42 \pm 3\%$ increase ($n=12$) and slices treated with 2 μ M BN-52021 alone exhibited a $2 \pm 2\%$ increase ($n=10$).

BN-52021 also inhibited LTP when administered for 5 min immediately after the tetanus (Fig. 1e). Slices exposed to post-tetanic perfusion of 5 μ M BN-52021 exhibited a $9 \pm 3\%$ increase in e.p.s.p. slope 20–30 min after the tetanus ($n=9$), compared with a $35 \pm 3\%$ increase in slices exposed to bath solution applied in a similar fashion ($n=7$). In all experiments, the effects of BN-52021 were reversible, with a second tetanus administered 20–40 min after drug washout resulting in LTP.

These observations indicate that specific PAF receptors contribute to CA1 LTP and are consistent with previous studies using less selective PAF antagonists^{11,12}. Furthermore, in cultured rat hippocampal neurons, MC-PAF, acting at BN-52021-sensitive receptors, selectively augments excitatory postsynaptic currents (e.p.s.cs) by a presynaptic mechanism⁸. To determine whether

this presynaptic enhancement could result from PAF generated in postsynaptic neurons, we examined the effects of postsynaptic applications of MC-PAF on miniature e.p.s.cs. When included in the pipette solution during whole-cell recordings, 2 μ M MC-PAF produced an increase in the frequency, but not the amplitude, of miniature e.p.s.cs (Fig. 2a–d). The increase in miniature e.p.s.c. frequency occurred about 6 min after intracellular contact had been established and persisted for 6–10 min, subsequently returning to baseline. The effects of MC-PAF were completely inhibited by 2 μ M BN-52021 administered extracellularly. To examine whether the effects of MC-PAF resulted from extracellular drug leakage, in three neurons we included 10 μ M MC-PAF in the intracellular solution and delayed whole-cell penetration for 10 min after the formation of an on-cell seal. In these cells, the increase in miniature e.p.s.c. frequency required 5 min, similar to the delay observed with 2 μ M MC-PAF (Fig. 2e). In these cells, the increase in miniature e.p.s.c. frequency persisted for at least 20 min, suggesting that postsynaptically generated PAF can have longer-term effects on synaptic transmission.

The apparent presynaptic effects of postsynaptically applied MC-PAF are consistent with PAF or a PAF-derived messenger serving as a retrograde signal in CA1 LTP. However, the effects of 2 μ M MC-PAF are time-limited, suggesting that, at low concentrations, PAF may require other factors to produce LTP. To

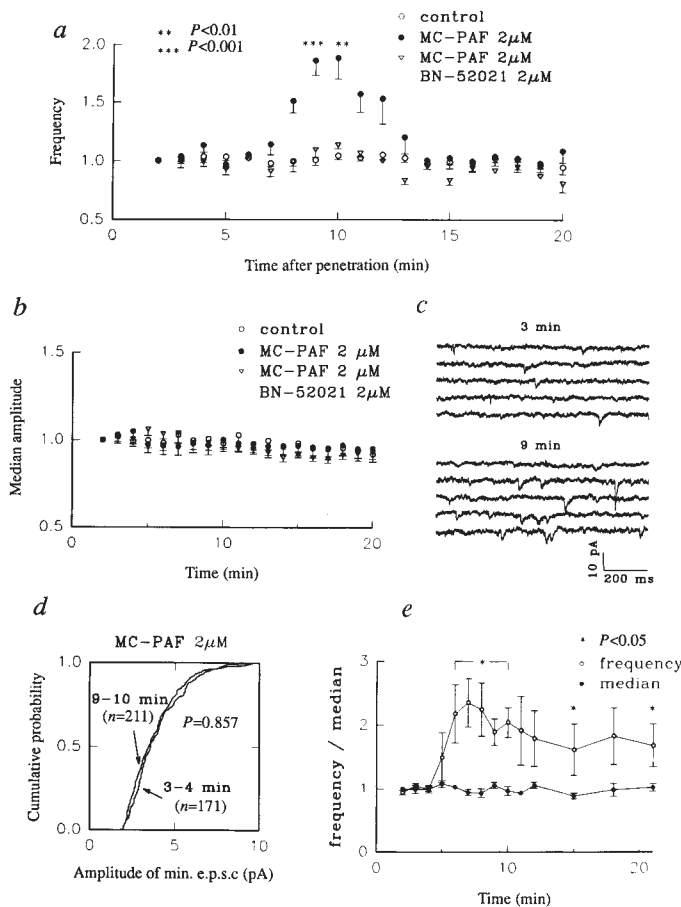
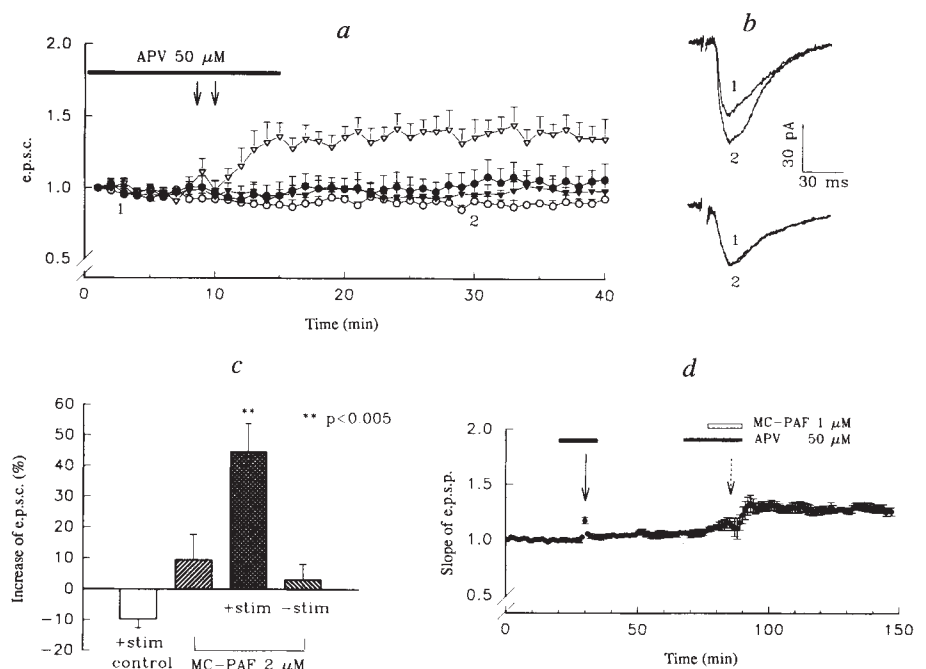


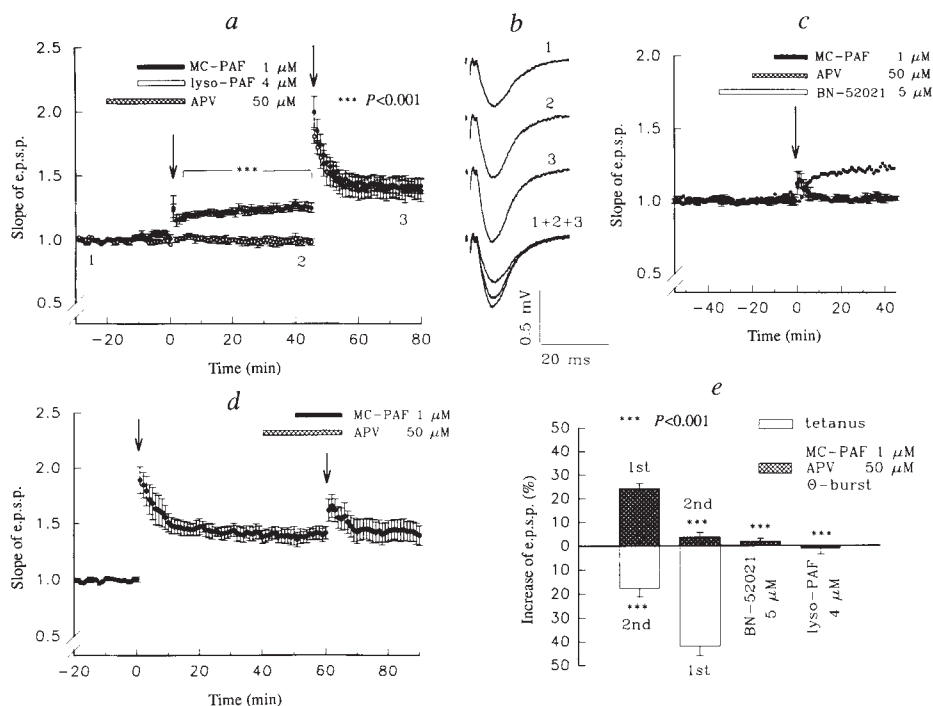
FIG. 2 Postsynaptically applied MC-PAF produced a time-dependent increase in the frequency but not the amplitude of miniature e.p.s.cs. **a**, The graph shows the normalized average frequency of miniature e.p.s.cs obtained from neurons recorded using control intracellular solution (open circles; $n = 6$) and intracellular solution containing 2 μ M MC-PAF (filled circles; $n = 8$). BN-52021 (2 μ M) applied in the extracellular solution blocked the increase in miniature e.p.s.c. frequency (open triangles; $n = 4$). $**P < 0.01$; $***P < 0.001$ compared with control by Mann-Whitney U -test. **b**, The graph shows the normalized median amplitude of miniature e.p.s.cs obtained under the conditions depicted in **a**. Median values are shown because miniature e.p.s.c. amplitudes are not normally distributed⁸. **c**, The traces depict consecutive miniature e.p.s.cs before and during the period that MC-PAF enhanced the frequency. **d**, The graph shows representative cumulative probability plots of miniature e.p.s.c. amplitudes larger than 2 pA obtained 3–4 min and 9–10 min after penetration into the neuron shown in **c**. The 9–10-min curve represents the period of peak miniature e.p.s.c. frequency increase. To determine whether the amplitude distribution of the two populations differed, the nonparametric Kolmogorov-Smirnov test was used. **e**, The graph shows the increase in miniature e.p.s.c. frequency (open circles) and lack of change in median miniature e.p.s.c. amplitude (solid circles) in three neurons in which 10 μ M MC-PAF was included in the intracellular solution. In these cells, entry into the whole-cell mode was delayed for 10 min after the formation of an on-cell seal. **METHODS**. In experiments examining effect on miniature e.p.s.cs, 1 μ M tetrodotoxin (TTX) and 100 μ M PTX were added to the extracellular solution and neurons were voltage clamped at -90 mV. Data were recorded on tape, filtered at 1 kHz, digitized at 5 kHz and analysed using a routine written in Axobasic. Events (≥ 2 pA) were detected by setting a threshold that excluded baseline noise and were verified by visual inspection at the time of analysis to exclude electrical artefacts or changes in the baseline. Events that occurred during the decay of a prior miniature e.p.s.c. were not analysed for peak current but were included in frequency determinations.

FIG. 3 Presynaptic activation plus MC-PAF produces LTP. **a**, The filled circles show the evoked e.p.s.c. amplitude when 2 μ M MC-PAF was included in the intracellular solution during whole-cell recording in neurons voltage clamped at -80 mV ($n = 5$). Open triangles show intracellular applications of 2 μ M MC-PAF combined with 1 Hz stimulation of the Schaffer collateral pathway for 15 s (arrows) ($n = 8$). Two stimulations, each of 1 Hz for 15 s, were administered 2 min apart. In four of the eight slices examined using postsynaptic MC-PAF and 1 Hz stimulation, e.p.s.cs from a non-stimulated pathway were also examined and failed to show LTP (filled triangles). Open circles show control e.p.s.cs from neurons recorded in the absence of MC-PAF and not given 1 Hz stimulation ($n = 5$). $\square$ -APV (50 μ M) was administered for the period denoted by the solid bar. Results are normalized with respect to the first e.p.s.c.s obtained after entry into the neuron. **b**, The traces depict e.p.s.cs taken at the times indicated by the numbers in **a** from the 1-Hz stimulated pathway (top traces) and the unstimulated pathway (bottom traces) in a single neuron. **c**, The bar graph shows the per cent change in e.p.s.c. amplitude measured 20–30 min after 1 Hz stimulation. **d**, In separate slices, bath applications of 1 μ M MC-PAF produced LTP when combined with 1 Hz stimulation for 20 s (dashed arrow) in the presence of 50 μ M $\square$ -APV. In these slices, 50 μ M $\square$ -APV prevented θ -burst-induced LTP



(solid arrow). In the absence of APV, θ -bursts (5 pulses at 100 Hz every 200 ms for 1 s) produced a $+44 \pm 3.7\%$ change in e.p.s.p. slope ($n = 5$).

FIG. 4 Bath applications of MC-PAF plus 0-burst stimulation in the presence of APV produce LTP and diminish tetanus-induced LTP. **a**, The graph shows the time course of change in field e.p.s.p. slope for slices treated with 1 μ M MC-PAF (solid symbols; $n=8$) or 4 μ M lyso-PAF (open symbols; $n=4$). At the time denoted by the first arrow a 0-burst stimulus was applied. The bars above the graph denote the application of MC-PAF, lyso-PAF and 50 μ M APV. At the second arrow a 100 Hz $\times$ 1 s tetanus was administered. **b**, The traces show e.p.s.ps taken during the times denoted by the numbers in **a** in a slice treated with MC-PAF. **c**, BN-52021 (5 μ M) blocked LTP induced by MC-PAF plus 0-burst stimulation (filled symbols; $n=5$). The open circles show the time course of PAF-mediated LTP obtained by subtracting the BN-52021 results from the data shown in the filled symbols of **a**. **d**, Tetanus-induced LTP completely occluded PAF-induced potentiation. The graph shows the change in e.p.s.p. slope for slices administered 100 Hz $\times$ 1 s tetanic stimulation (first arrow) followed by 1 μ M MC-PAF, 50 μ M APV and 0-burst stimulation ($n=5$). **e**, The bar graph shows the degree of enhancement produced by MC-PAF plus 0-burst stimulation compared with tetanic stimulation when MC-PAF is applied before (1st) or after (2nd) tetanic stimulation. The last



two bars show the effects of BN-52021 and lyso-PAF. Statistical differences were determined by comparing the two MC-PAF applications or the two tetanic stimulations. BN-52021 and lyso-PAF were compared with the first MC-PAF administration.

examine this issue further, we combined postsynaptic administration of 2 μ M MC-PAF with weak presynaptic stimulation (two stimulations each of 1 Hz for 15 s, 2 min apart) in the presence of 50 μ M D-2-amino-5-phosphonovalerate (APV), an NMDA-receptor antagonist. For these experiments, postsynaptic neurons were voltage-clamped at -80 mV throughout the experiment. In control neurons, 1 Hz stimulation failed to induce LTP ($-9.7 \pm 2.8\%$ change in e.p.s.c. amplitude; $n=5$) (Fig. 3a-c). However, when 2 μ M MC-PAF was present in the postsynaptic cell and 1 Hz stimulation was administered 8–10 min after whole-cell penetration, a significant and persisting enhancement of e.p.s.c.s was observed ($+44.5 \pm 9.3\%$ change; $n=8$). The LTP was synapse-specific in that a non-stimulated input to CA1 failed to show enhancement ($+3.2 \pm 5.0\%$ change; $n=4$). MC-PAF (2 μ M) administered postsynaptically in the absence of 1 Hz stimulation failed to produce LTP ($+9.4 \pm 8.3\%$ change; $n=5$).

These observations suggest that PAF in combination with weak presynaptic stimulation is sufficient to produce LTP in the presence of APV. Similarly, we found that bath applications of 1 μ M MC-PAF produced LTP of field e.p.s.ps when combined with 1 Hz (Fig. 3d) or 0-burst stimulation (Fig. 4a, b) in the presence of 50 μ M APV ($+27.2 \pm 3.2\%$ ($n=6$) and $+24.2 \pm 2.4\%$ ($n=9$) change in e.p.s.p. slope, respectively). Control slices given 0-burst stimulation in the presence of APV failed to exhibit LTP ($+4.1 \pm 1.7\%$ change, $n=7$). The effects of MC-PAF were not mimicked by 4 μ M lyso-PAF, an inactive analogue ($-1.1 \pm 2.4\%$ change, $n=4$) (Fig. 4a) and were blocked by 5 μ M BN-52021 ($+1.7 \pm 1.3\%$ change, $n=5$) (Fig. 4c). The enhancement produced by the combination of MC-PAF and 0-burst stimulation significantly diminished the potentiation produced by a strong tetanus (Fig. 4a, e). Similarly, LTP induced by a strong tetanus completely occluded PAF-induced potentiation (Fig. 4d, e).

The dependence of CA1 LTP on NMDA-receptor activation and postsynaptic increases in intracellular Ca^{2+} , coupled with evidence supporting presynaptic changes in LTP^{1,2}, has

prompted the hypothesis that a Ca^{2+} -dependent trans-synaptic messenger is responsible for the augmented release of transmitter. The leading candidates for such a messenger are nitric oxide (NO), carbon monoxide (CO) and arachidonic acid⁴. In the case of NO, postsynaptic applications of NO synthase inhibitors block LTP^{6,13}, as does extracellular haemoglobin, an agent that binds NO (refs 6, 13). Furthermore, NO increases the frequency of miniature e.p.s.c.s in cultured hippocampal neurons, suggesting that the agent acts by a presynaptic mechanism⁶. NO and CO produce CA1 LTP only when administered in conjunction with presynaptic activation¹⁴. The source of NO remains unclear, as CA1 pyramidal neurons do not seem to contain NO synthase¹⁵. It is possible that other cells in the CA1 region release NO or that pyramidal neurons contain a novel version of the enzyme. Other studies suggest that NO contributes to LTP only under certain conditions, including intense synaptic stimulation¹⁶ and low temperature¹⁷, and that untimely release of NO inhibits, rather than promotes, LTP¹⁸.

Several studies suggest that phospholipase A₂ products are important in LTP^{1,19}. Phospholipase A₂ promotes formation of both arachidonic acid and PAF^{7,20}, making PAF a candidate for mediating effects attributed to phospholipase A₂. Both arachidonic acid⁵ and PAF produce LTP when paired with presynaptic activation. However, the effects of arachidonic acid develop slowly, require high concentrations⁵, and, in CA1, are inhibited by APV^{6,21}. Prolonged bath applications of PAF also produce a slowly developing potentiation that is blocked by NMDA-receptor antagonists²². As arachidonic acid potentiates NMDA responses²³ and PAF promotes the release of arachidonic acid^{18,24}, these slowly developing forms of LTP could result solely from the effects of arachidonic acid. Our results indicate that, when coupled with weak presynaptic activation or possibly when applied at higher concentrations alone, PAF produces LTP that does not require NMDA receptors. This suggests that PAF acts downstream of NMDA receptors in the LTP cascade. PAF-

induced LTP significantly diminishes tetanus-induced LTP and tetanic LTP completely occludes PAF potentiation, suggesting that the two forms of synaptic enhancement share common mechanisms. The failure of PAF-mediated potentiation to occlude tetanus-induced LTP completely suggests that tetanus-induced LTP contains a PAF-insensitive component.

The observation that PAF is released extracellularly from cultured neurons²⁵ is consistent with a proposed intercellular messenger role^{26,27}, and recent evidence indicates that there are functional cell-surface PAF receptors in the central nervous system which increase intracellular Ca^{2+} (ref. 28). Our studies indicate that PAF serves as a component in the biochemical events that modify the release of excitatory neurotransmitter during LTP²⁹. It is unclear whether PAF or a PAF-derived messenger is the sole mediator of the presynaptic effects. □

Received 28 July; accepted 1 November 1993.

1. Kuba, K. & Kumamoto, E. *Prog. Neurobiol.* **34**, 197–269 (1990).
2. Bekkers, J. M. & Stevens, C. F. *Nature* **346**, 724–728 (1990).
3. Manabe, T., Renner, P. & Nicoll, R. A. *Nature* **355**, 50–55 (1992).
4. Kandel, E. R. & O'Dell, T. J. *Science* **258**, 243–245 (1992).
5. Williams, J. H., Errington, M. L., Lynch, M. A. & Bliss, T. V. P. *Nature* **341**, 739–742 (1989).
6. O'Dell, T. J., Hawkins, R. D., Kandel, E. R. & Arancio, O. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11285–11289 (1991).

7. Bazan, N. G. *Ann. N.Y. Acad. Sci.* **559**, 1–16 (1989).
8. Clark, G. D., Happel, L. T., Zorumski, C. F. & Bazan, N. G. *Neuron* **9**, 1211–1216 (1992).
9. Marcheselli, V. L., Rossowska, M. J., Domingo, M. T., Braquet, P. & Bazan, N. G. *J. biol. Chem.* **265**, 9140–9145 (1990).
10. Marcheselli, V. L. & Bazan, N. G. *J. Neurosci. Res.* (in the press).
11. Del Cerro, S. D., Arai, A. & Lynch, G. *Behav. neural Biol.* **54**, 213–217 (1990).
12. Arai, A. & Lynch, G. *Eur. J. Neurosci.* **4**, 411–419 (1992).
13. Schuman, E. M. & Madison, D. V. *Science* **254**, 1503–1506 (1991).
14. Zhuo, M., Small, S. A., Kandel, E. R. & Hawkins, R. D. *Science* **260**, 1946–1950 (1993).
15. Bredt, D. S. et al. *Neuron* **7**, 615–624 (1991).
16. Gribkoff, V. K. & Lum-Ragan, J. T. *J. Neurophysiol.* **68**, 639–642 (1992).
17. Li, Y. G., Errington, M. L., Williams, J. H. & Bliss, T. V. P. *Soc. Neurosci. Abstr.* **18**, 343 (1992).
18. Izumi, Y., Clifford, D. B. & Zorumski, C. F. *Science* **257**, 1273–1276 (1992).
19. Lynch, M. A., Errington, M. L. & Bliss, T. V. P. *Neuroscience* **30**, 693–701 (1989).
20. Braquet, P., Touqui, L., Shen, T. Y. & Vargaftig, B. B. *Pharmac. Rev.* **39**, 97–145 (1987).
21. Kato, K., Urano, K., Saito, K. & Kato, H. *Brain Res.* **563**, 94–100 (1991).
22. Wieraszko, A., Li, G., Kornecki, E., Hogan, M. V. & Ehrlich, Y. H. *Neuron* **10**, 553–557 (1993).
23. Miller, B., Sarantis, M., Traynelis, S. F. & Attwell, D. *Nature* **355**, 722–725 (1992).
24. Doucet, J. P. & Bazan, N. G. *Molec. Neurobiol.* **6**, 407–424 (1992).
25. Yue, T.-L., Lysko, P. G. & Feuerstein, G. J. *Neurochem.* **54**, 1809–1811 (1990).
26. Prescott, S. M., Zimmerman, G. A. & McIntyre, T. M. *J. biol. Chem.* **265**, 17381–17384 (1990).
27. Bazan, N. G. & Cluzel, J. M. in *Emerging Strategies in Neuroprotection* (eds Marangos, P. J., & Lal, H.) 238–251 (Birkhauser, Boston, 1992).
28. Bito, H. et al. *Neuron* **9**, 285–294 (1992).
29. Lynch, M. A. & Voss, K. L. *J. Neurochem.* **56**, 113–118 (1991).
30. Kato, K., Clifford, D. B. & Zorumski, C. F. *Neuroscience* **53**, 39–47 (1993).

ACKNOWLEDGEMENTS. This work was supported by grants from the NIMH, NIH, NIA and the Bantky Foundation. We thank A. Benz and J. Que for technical assistance and Y. Izumi, D. B. Clifford and S. Mennerick for discussion.

Evidence that the S6 segment of the Shaker voltage-gated K^+ channel comprises part of the pore

George A. Lopez, Yuh Nung Jan & Lily Y. Jan

Howard Hughes Medical Institute and Departments of Physiology and Biochemistry, University of California at San Francisco, Parnassus and Third Avenues, U-426, San Francisco, California 94143-0724, USA

POTASSIUM channels are highly selective and allow the rapid flux of potassium ions through their pore¹. Several studies have implicated the H5 (P or SS1-SS2) segment^{2,3} as part of the pore in voltage-gated ion channels^{4–10}. The proposal that H5 spans at least 80% of the electric potential drop across the K^+ channel pore^{5,11} is based on altered internal tetraethylammonium sensitivity arising from mutations of H5 residues that are 100% conserved among K^+ channels having differing sensitivity to tetraethylammonium^{5–7,12,13}. Here we report that the S6 segment is also involved in K^+ ion permeation and in governing the sensitivity to internal tetraethylammonium and barium. Transplanting the S6 segment of NGK2 into Shaker causes this S6 chimaera to adopt the single-channel conductance and sensitivity to internal tetraethylammonium and barium ions from the NGK2 channel. The differences between NGK2 and Shaker in external tetraethylammonium sensitivity, but not single-channel conductance, can be attributed to the differences in their H5 sequences. Three non-conserved S6 residues have been found to affect either single-channel conductance or internal tetraethylammonium sensitivity.

To determine whether the S6 segment forms part of the permeation pathway, we constructed chimaeric channels between Shaker B (ref. 14) and NGK2 (ref. 15), and measured several pore properties that differ between these two potassium channels, which are from different subfamilies¹. The NGK2 channel has a single-channel conductance larger than that of Shaker B^{5,6}. In addition, the NGK2 channel is approximately two orders of magnitude more sensitive to external block by tetraethylammonium (TEA), and over an order of magnitude less sensitive to internal block by TEA than Shaker B^{5,6}. The S6 chimaera was

constructed by transplanting an S6-containing segment of NGK2 to replace that of Shaker B (Fig. 1), which contains a deletion of amino acid residues 6–46 (ShBΔ). The amino-terminal deletion removes fast inactivation which would otherwise interfere with the action of open-channel blockers such as TEA and barium ions, but does not affect potassium permeation¹⁶. If the S6 segment forms part of the permeation pathway, then we would expect several of the pore properties of the S6 chimaera to resemble those of NGK2 rather than ShBΔ.

Transplanting the S6 region (Fig. 1) from NGK2 into ShBΔ increased the single-channel conductance by fourfold (Fig. 1 and

TABLE 1 Single-channel conductances and internal TEA sensitivities from the Shaker mutants

Mutant	Chord conductance (pS) (mean ± s.d. (N))	K_i^* (mM) (mean ± s.d. (N))
ShBΔ	11.2 ± 1.0 (4)	0.23 ± 0.06 (18)
S6 chimaera	43.8 ± 1.7 (5)	5 ± 1.3 (17)
NGK2	27.0 ± 1.0 (7)	8 ± 2.2 (6)
H5 chimaera	11.6 ± 0.3 (3)	ND†
I464L, L472M	12.6 ± 0.8 (3)	4.8 ± 1.4 (5)
S479N, N482G, Y483M	20.5 ± 1.5 (3)	0.34 ± 0.06 (3)
F484Y, H486S, R487L	22.8 ± 1.1 (5)	0.17 ± 0.04 (4)
I464L	ND	0.16 ± 0.02 (5)
L472M	ND	3.1 ± 1.3 (4)
S479N	16.7 ± 0.6 (3)	ND
N482G	14.4 ± 0.6 (3)	ND
Y483M	12.6 ± 0.7 (4)	ND
F484Y	11.2 ± 0.3 (3)	ND
H486S	17.6 ± 1.2 (5)	ND
R487L	13.7 ± 1.4 (4)	ND

Excised inside-out patches were used with the recording solutions used in Fig. 1. Mutations were constructed by cassette mutagenesis as described in Fig. 1 legend. Chord conductances were determined as described for Fig. 1. Amino acids are designated in the one-letter code. ND, not determined.

* K_i s, or half-blocking concentrations for internally applied TEA at 0 mV, were determined by fitting the fractional current remaining after an 80-ms voltage pulse to an equation that describes open-channel block at a single site²⁸, except for ShBΔ, NGK2 and the S6 chimaera, which were determined from the TEA dose-response curves of Fig. 2.

† Not determined because of low expression.

FIG. 1 The ShBA-NGK2 S6 chimaeric potassium channel and single-channel properties of ShBA, NGK2 and the S6 chimaera. *a*, Amino-acid alignment of the sixth proposed transmembrane segment (S6) and adjacent residues of the Shaker B potassium channel¹⁴ and of the mammalian NGK2 channel¹⁵. Boxed region represents the proposed sixth transmembrane segment. The dashed bracket corresponds to the region that was transplanted from the NGK2 channel into ShBA. Dashes represent residues of NGK2 that are identical to the corresponding residues in ShBA. The numbers apply to the first residue in each of the aligned sequences. N and C, amino and carboxy termini. *b*, Representative single-channel currents recorded from excised inside-out patches from *Xenopus* oocytes. *c*, Representative open channel current-voltage relations. V_m , membrane potential, I , single-channel current.

METHODS. Restriction sites that flank the S6 segment but do not alter the amino-acid sequence were introduced into ShBA. The two restriction sites *Nru*I and *Hind*III were introduced by oligonucleotide-directed mutagenesis as described²⁹. The mutated DNA was then cut with the two restriction enzymes and gel-purified. For cassette mutagenesis, a double-stranded DNA segment was constructed with synthetic oligonucleotides containing the coding sequence of the corresponding region from the NGK2 clone, ligated into the ShBA backbone and verified by sequencing. To simplify the kinetic analysis and measurements, we used an amino-terminal deletion mutant of Shaker B, ShBA, which does not show fast inactivation¹⁶. The H5 chimaera was similarly constructed by cassette mutagenesis using the restriction sites *Nsi*I and *Hind*III which flank the H5 region of ShBA (amino acids 433–460). The corresponding H5 sequence of NGK2 (ref. 15), identical to that transplanted into DRK1 in a previous study⁶, was used to replace the H5 of ShBA. Single-channel currents were recorded from excised inside-out patches using a List EPC-7 amplifier, sampled at 5 kHz and filtered at 1 kHz. The external (pipette) solution contained 98 mM KCl, 1 mM MgCl₂, 0.3 mM CaCl₂, 10 mM HEPES (to pH 7.1 with KOH). The internal (bath) solution contained 98 mM KCl, 0.5 mM MgCl₂, 1 mM EGTA and 10 mM HEPES (to pH 7.1 with KOH) (20–23 °C). Holding potential was –100 mV and the membrane was depolarized to +100 mV for 320 ms. Scale bar, 1.4 pA, 30 ms. Single-channel current amplitudes were measured from amplitude histograms and were leak-subtracted using blank traces that contained no channel openings.

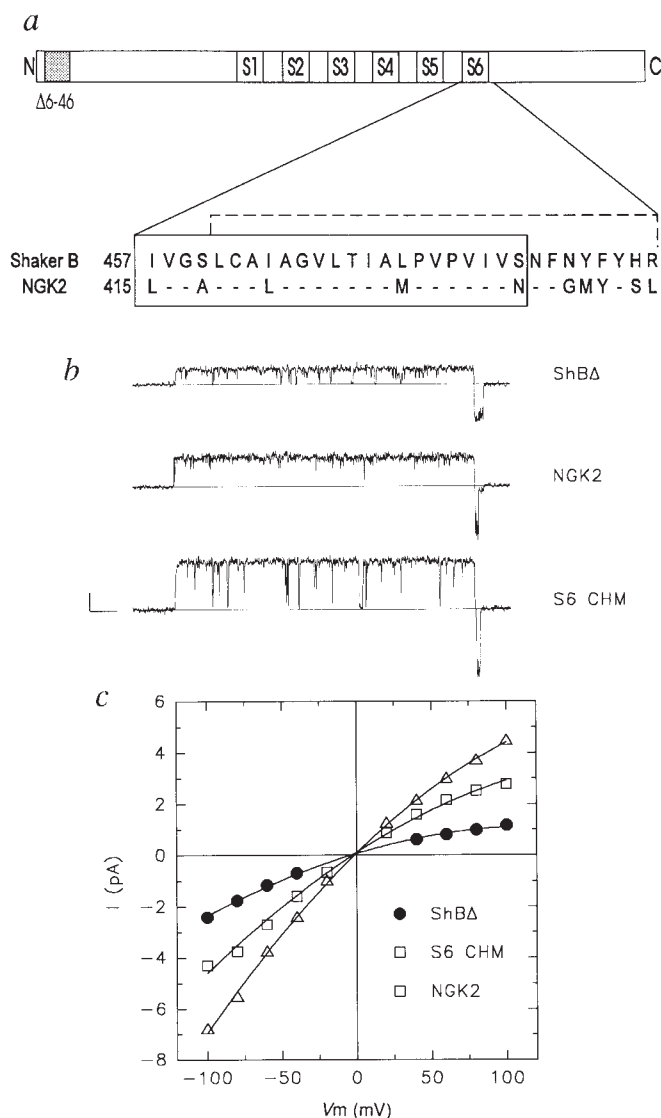


Table 1), from 11 pS to 44 pS. This increase in single-channel conductance was unexpected, given that the difference in the single-channel conductance between NGK2 (27 pS) and DRK1 (8 pS) had been attributed to the H5 sequence in a similar analysis of chimaeric potassium channels⁶. The identical H5 segment of NGK2, which confers a large conductance when transplanted into DRK1 (ref. 6), did not alter the single-channel conductance when transplanted into ShBA, even though this H5 chimaera showed an external TEA sensitivity ($K_i = 0.6$ mM) similar to that of NGK2 ($K_i = 0.3$ mM) rather than to that of ShBA ($K_i = 21$ mM) or the S6 chimaera ($K_i = 23$ mM) (Figs 2a and 3; Table 1). Thus, both the H5 and S6 segment appear to limit the conductance, perhaps by forming a tight constriction, but only the H5 segment contains the external binding site for TEA. The small conductance of DRK1 is probably due to its H5 segment restricting potassium flux, whereas the small conductance of ShBA appears to rise primarily from a limiting S6 segment.

The S6 chimaera not only increased the single-channel conductance but it also altered the sensitivity to internally applied TEA and barium ions. Whereas TEA blocks ion conduction by occluding the mouth of the pore^{17,21}, barium ions block the potassium channel pore by interacting with potassium binding sites at various depths in the pore^{1,22,23}, probably as a result of the similar crystal radii of barium and potassium ions (1.35 Å and 1.33 Å, respectively¹). The presence of 50 μM internal bar-

ium ions produced a time- and voltage-dependent block of the NGK2 potassium channels (Fig. 2g, h) owing to the relatively slow rate of barium block of the open channel^{22,23}. Compared to the block of NGK2 channels, the barium block of ShBA channels was much slower and less effective (Fig. 2c, d). The barium block of the S6 chimaera was more similar to that of NGK2 than ShBA (Fig. 2e, f). In particular, the barium block showed much greater voltage dependence for NGK2 and the chimaeric channel than for ShBA (Fig. 2c–i). Thus, barium appears to penetrate further into the pore of the NGK2 and chimaeric channels, so that its binding to the channel pore is influenced by the electric potential across the membrane. This suggests that transplanting the S6-containing segment of NGK2 into ShBA has either introduced a barium binding site in the pore or has exposed such a site by reducing an energy barrier to barium entry. In addition, the internal TEA sensitivity of the S6 chimaera was similar to that of NGK2 rather than that of ShBA (Fig. 2b; Table 1), suggesting that the internal mouth or vestibule of the pore of the S6 chimaera resembles that of NGK2.

The S6 transplantation did not alter the voltage dependence or kinetic properties of channel gating (see legend to Fig. 3), suggesting that the overall channel structure was not affected. Given that the differences in the H5 sequence between NGK2 and ShBA accounted largely for the differences in external TEA sensitivity but not single-channel conductance, it seems unlikely

that the effect of the S6 transplantation on single-channel conductance could be attributed to an indirect effect via H5.

Eight amino acids differ in ShBΔ and the S6 chimaera. To determine whether the alterations in pore properties arise from one or multiple amino-acid differences in the S6 region, we analysed three 'mini-chimaeras': I464L/L472M, S479N/N482G/Y483M and F484Y/H486S/R487L (single-letter amino-acid notation; numbers indicate residue position). The two triple mutations had single-channel conductance (Fig. 3; Table 1) and internal barium sensitivity (E. Isacoff, personal communication) intermediate between those of ShBΔ and NGK2, but their internal TEA sensitivity remained very similar to that of ShBΔ (Table 1). In contrast, the double mutant had a single-channel conductance identical to that of ShBΔ although its internal TEA sensitivity was the same as that of the S6 chimaera (Fig. 3; Table 1). Of the two mutations included in the double mutant, the L472M mutation, but not the I464L mutation, had a reduced internal

TEA sensitivity similar to that of the double mutant (Table 1). Among the six mutations included in the two triple mutants with a larger single-channel conductance, S479N and H486S increased the single-channel conductance (Fig. 3; Table 1). The mutation A463V alters the rate of slow inactivation and single-channel conductance²⁴, and mutations of T469 alter the sensitivity to internally applied quaternary ammonium blockers with extended alkyl chains without affecting the sensitivity to internal TEA²⁵. Thus, multiple residues spanning the S6 segment have been found to affect pore properties.

The L472M mutation alone accounted for the effect of the S6 chimaera on internal TEA sensitivity. Mutations of M440 and T441 in H5 are also known to alter internal TEA sensitivity^{5,25}. But M440 and T441 are 100% conserved among all known voltage-gated potassium channel sequences and therefore cannot account for the different internal TEA sensitivities of these channels. In contrast, there is a leucine residue at position 472 in the

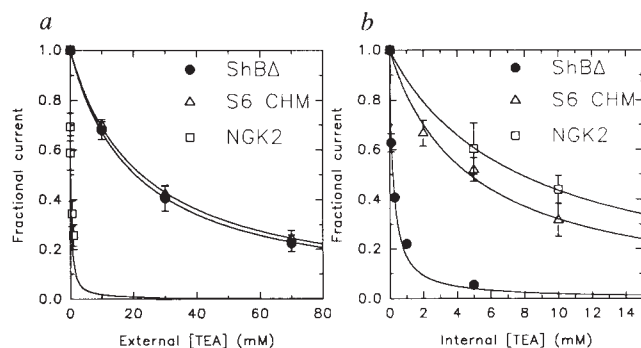


FIG. 2 TEA and barium ion blockade of currents from ShBΔ, NGK2 and the S6 chimaeric channel. *a*, Sensitivity to external TEA. The reduction in peak current at 0 mV for ShBΔ and the chimaera, or at +20 mV for NGK2, was measured as a function of external TEA concentration. *b*, Sensitivity to internal TEA. The reduction in peak current at 0 mV was measured as a function of TEA concentration. Each point is the mean from at least six determinations; error bars indicate the s.d. The curves were fitted to the equation: fractional current = $1/(1 + [TEA]/K_i)$. *c-i*, Sensitivity to internal barium ions. Ionic currents from ShBΔ (*c,d*), the S6 chimaera (*e,f*) and NGK2 (*g,h*), before (*c,e,g*) and after (*d,f,h*) application of internal barium ions. Excised inside-out patches were exposed to the same internal solution as in Fig. 1 (*c,e,g*) before perfusion with solution that was of the same composition except that 50 μM barium chloride was added (*d,f,h*). Holding potential was -100 mV with 160-ms depolarizing pulses from 20 mV to 100 mV in 40 mV increments for *c*, *e* and *g*, or from 20 mV to 140 mV in 40 mV increments for *d*, *f* and *h*. Scale bar: 980 pA, 32 ms. *i*, Voltage dependence of current block. The fast component of current block was fitted with a single exponential and the time constant was plotted as a function of membrane potential. The points are the means from at least three determinations and the error bars indicate s.d. The line is a first-order linear regression through the data points.

METHODS. Experiments for the external TEA sensitivity were performed on RNA-injected, voltage-clamped *Xenopus* oocytes using a two-electrode voltage clamp as described³⁰, at a holding potential of -100 mV. External solutions contained: TEA-Cl at concentrations indicated, varying concentrations of NaCl to maintain the total concentration of NaCl and TEA-Cl at 88 mM, 1 mM KCl, 2.4 mM NaHCO₃, 0.82 mM MgSO₄, 0.33 mM Ca(NO₃)₂, 0.41 mM CaCl₂ and 10 mM HEPES (to pH 7.5 with NaOH). Each point is the mean from at least five determinations and error bars indicate the s.d. The curves were fitted to the equation: fractional current = $1/(1 + [TEA]/K_i)$, with the K_i values for external TEA: $K_i = 21 \pm 3.6$ mM (10), 23 ± 2.5 mM (7), and 0.2 ± 0.5 mM (5), mean $\pm$ s.d. (*N*) for ShBΔ, the S6 chimaeric channel, and NGK2. For the internal TEA or barium sensitivity, experiments were done on excised inside-out patches from a holding potential of -100 mV. TEA or barium was applied by perfusing the intracellular side of the patch with a perfusion pipette brought close to the membrane.

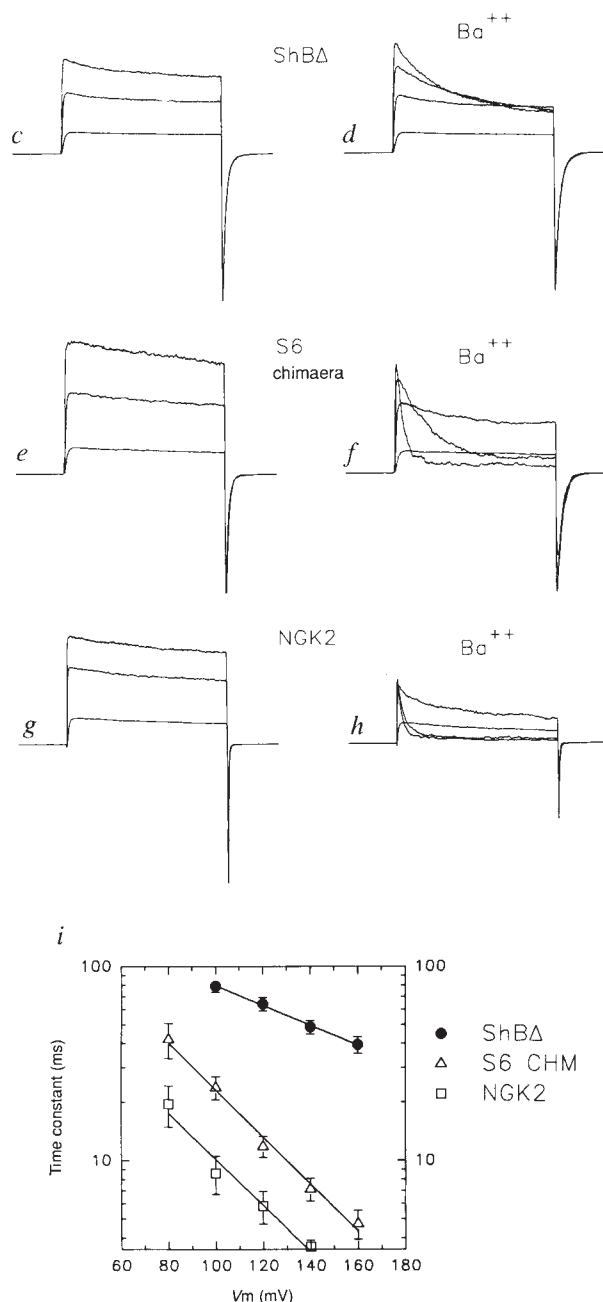
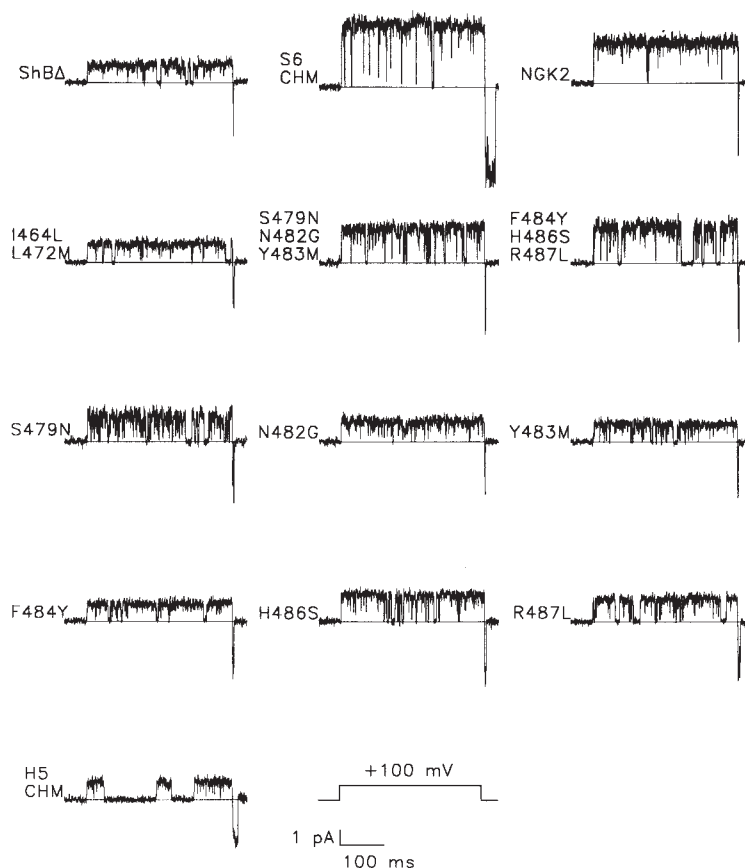


FIG. 3 Single-channel traces from the wild-type channels ShBA and NGK2, the H5 chimaera, the full S6 chimaera, the S6 'mini-chimaeras' and the individual point mutations in the S6 domain.

METHODS. The methods were as described for Fig. 1. To control for the possibility that the chimaeric mutation affected protein structure and altered channel function in a non-specific manner, we measured the voltage-dependence of activation and the maximum open-channel probability of the S6 chimaera with those of ShBA channels. The voltage-dependence of activation was measured from the size of macroscopic tail currents following activation voltage pulses to various voltages. The voltage-dependence of the normalized tail current was fitted to a Boltzmann distribution to determine the V_{mid} (mV) and slope (mV/e-fold change). Values for ShBA: $V_{mid} = -45 \pm 2$ mV, slope = 3.7 ± 0.7 mV/e-fold, $N = 8$; Values for the S6 chimaeric channel: $V_{mid} = -44 \pm 2$ mV, slope = 3.9 ± 0.4 mV/e-fold, $N = 3$ (mean $\pm$ s.d.). The tail currents at -100 mV were fitted with a single exponential using the program Clampfit (Axon Instruments). The time constants are not significantly different for ShBA: 3.4 ± 0.5 ms, $N = 5$, and the S6 chimaera: 3.2 ± 0.7 ms, $N = 4$, mean $\pm$ s.d. The open-channel probability was measured at $+100$ mV with each digitized point counted as either open or closed according to a 50% current threshold. P_{open} for ShBA: 0.88 ± 0.03 , $N = 4$; P_{open} for the S6 chimaeric channel: 0.92 ± 0.02 , $N = 4$ (mean $\pm$ s.d.).



Shaker subfamily, the members of which have a high sensitivity to internal TEA^{12,13}, whereas a methionine is found in the corresponding position in all known members of the Shaw subfamily including NGK2, which have a lower sensitivity to internal TEA^{6,12,13}. Thus the different internal TEA sensitivities of different potassium channels are probably due to a difference in the S6 segment at the position corresponding to L472 in Shaker

B. As differences between NGK2 and ShBA in the blockade of the pore by cytoplasmically applied TEA and barium ions, as well as the permeation rate of potassium ions, can be accounted for by differences in their S6 rather than H5 sequences, our results suggest that the pore of voltage-gated potassium channels is likely to be formed by the S6 segment, in addition to the H5 sequence^{5-7,10} and the S4-S5 intracellular loop^{26,27}. □

Received 10 September; accepted 22 October 1993.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1991).
- Guy, H. R. & Seetharamulu, P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 508–521 (1986).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 770–775 (1987).
- Yool, A. J. & Schwarz, T. S. *Nature* **349**, 700–704 (1991).
- Yellen, G., Jurman, M., Abramson, T. & MacKinnon, R. *Science* **251**, 939–942 (1991).
- Hartmann, H. A. et al. *Science* **251**, 942–944 (1991).
- Kirsch, G. E. et al. *Neuron* **8**, 499–505 (1992).
- Heinemann, S. H., Terlau, H., Stühmer, W., Imoto, K. & Numa, S. *Nature* **356**, 441–443 (1992).
- Backx, P. H., Yue, D. T., Lawrence, J. H., Marban, E. & Tomaselli, G. F. *Science* **257**, 248–251 (1992).
- Heginbotham, L., Abramson, T. & MacKinnon, R. *Science* **258**, 1152–1155 (1992).
- Miller, C. *Science* **252**, 1092–1096 (1991).
- Kirsch, G. E., Taglialatela, M. & Brown, A. M. *J. Physiol.* **261**, C583–590 (1991).
- Taglialatela, M. et al. *Molec. Pharmacol.* **40**, 299–307 (1991).
- Schwarz, T. S., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. *Nature* **331**, 137–142 (1988).
- Yokoyama, S. et al. *FEBS Lett.* **259**, 37–42 (1989).

- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. *Science* **250**, 533–538 (1990).
- Armstrong, C. M. & Binstock, L. J. *gen. Physiol.* **46**, 859–872 (1965).
- Armstrong, C. M. *J. gen. Physiol.* **50**, 491–503 (1966).
- Armstrong, C. M. *J. gen. Physiol.* **54**, 553–575 (1969).
- Armstrong, C. M. *J. gen. Physiol.* **58**, 413–437 (1971).
- Armstrong, C. M. & Hille, B. *J. gen. Physiol.* **59**, 388–400 (1972).
- Armstrong, C. M. & Taylor, S. R. *Biophys. J.* **30**, 473–488 (1980).
- Armstrong, C. M., Swenson, R. P. & Taylor, S. R. *J. gen. Physiol.* **80**, 663–682 (1982).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. *Neuron* **7**, 547–556 (1991).
- Choi, K. L., Mossman, C., Aube, J. & Yellen, G. *Neuron* **10**, 533–541 (1993).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. *Nature* **353**, 86–90 (1991).
- Slesinger, P. A., Jan, Y. N. & Jan, L. Y. *Neuron* **11**, 739–749 (1993).
- Woodhull, A. M. *J. gen. Physiol.* **61**, 687–708 (1973).
- Baldwin, T. J., Tsaour, M.-L., Lopez, G. A., Jan, Y. N. & Jan, L. Y. *Neuron* **7**, 471–483 (1991).
- Lopez, G. A., Jan, Y. N. & Jan, L. Y. *Neuron* **7**, 327–336 (1991).

ACKNOWLEDGEMENTS. We thank T. Baldwin, E. Isacoff, Y. Kubo, M. Li, E. Reuveny, M. Sheng, E. Shinbrot, P. Slesinger for reading the manuscript, M. Solis for technical assistance, S. Fried for DNA sequencing, E. Isacoff for the initial barium experiment, R. W. Aldrich for the 6–46 deletion mutation of Shaker B, and R. Joho and A. M. Brown for the NGK2 clone. Y.N.J. and L.Y.J. are Howard Hughes Investigators. This work was supported by the NIH.

High-affinity IgE receptor on eosinophils is involved in defence against parasites

Abdelillah Soussi Gounni*,
Bouchaib Lamkhied*, Kenichi Ochlaï†,
Yoichi Tanaka*, Emmanuel Delaporte‡,
André Capron*, Jean-Pierre Kinet†
& Monique Capron*§

* Centre d'Immunologie et de Biologie Parasitaire,
Unité Mixte INSERM U167-CNRS624, Institut Pasteur,
1 rue du Professeur A. Calmette, BP 245, 59019 Lille, France
† Molecular Allergy and Immunology Section, National Institute of
Allergy and Infectious Diseases, National Institutes of Health,
Rockville, Maryland 20852, USA
‡ Service de Dermatologie A, Centre Hospitalier Régional
Universitaire, 59000 Lille, France
§ To whom correspondence should be addressed

PARASITIC infections are often associated with eosinophilia and high levels of immunoglobulin E (IgE). This observation has led to speculation that eosinophils and IgE may act together in the immune response against parasites. In support of this hypothesis, IgE and eosinophils participate in cytotoxic reactions directed against *Schistosoma mansoni* larvae *in vitro*^{1,2}. Furthermore, epidemiological studies have shown an inverse correlation between levels of specific IgE and rates of infection with *Schistosoma*³⁻⁵. The low-affinity IgE receptor (FcεRII/CD23) was first incriminated in eosinophil activation^{6,7}. The fact that the high-affinity IgE receptor (FcεRI)^{8,9} is not only expressed on mast cells and basophils but also on Langerhans cells^{10,11} led us to investigate the presence of FcεRI on eosinophils. Here we show that FcεRI is expressed on eosinophils from hypereosinophilic patients, is involved in eosinophil degranulation, and participates in eosinophil-mediated cytotoxicity against *S. mansoni*. Our results indicate that FcεRI may play a major part in immune defence against parasites.

Highly purified eosinophils from hypereosinophilic patients were incubated with ¹²⁵I-labelled monomeric human IgE together with increasing concentrations of the anti-FcεRI α-chain monoclonal antibody 15-1 (ref. 10) or with a control antibody, and the cell-bound radioactivity counted. The monoclonal antibody 15-1, but not the isotopic control antibody, induces a

dose-dependent inhibition of monomeric IgE binding (Fig. 1). At the higher dose of 15-1 (50 µg ml⁻¹, inhibition is comparable to the inhibition seen with 1 mg ml⁻¹ of unlabelled human IgE. Under the same conditions, 1 mg ml⁻¹ of human IgG did not inhibit IgE binding, whereas 15-1 was not able to inhibit the binding of ¹²⁵I-labelled human IgG (data not shown). We then used flow cytometry to confirm the surface expression of FcεRI α-chain on the eosinophils. The FcεRI⁺ human basophil cell line KU812 served as a positive control. As expected, 90% of the KU812 cells stain positively with 15-1 (Fig. 2a) but not with control antibody. Under the same conditions, positive staining is detected on 13–73% of cells in purified eosinophil preparations (Fig. 2b, c). This percentage varies according to individual donors and is equivalent to binding of IgE (Fig. 2c). Immunostaining experiments with 15-1 were also performed on cyto-centrifuged KU812 cells or purified blood eosinophil preparations. Positive staining is again found on KU812 cells (Fig. 2d), as well as on some eosinophils (Fig. 2f). In contrast, no staining is detectable with the control antibody (Fig. 2e–g). Moreover, combined immunostaining of tissue eosinophils with 15-1 and an antibody against eosinophil cationic protein (ECP) shows that the 15-1-positive cells are eosinophils (Fig. 2h, i). Taken together, these data demonstrate that eosinophils from these patients can bind monomeric IgE through the FcεRI α-chain and that expression of FcεRI varies between individuals.

We next investigated whether transcripts for the FcεRI α-chain^{12,13}, which contains the binding site for IgE^{14,15}, could be detected by northern blotting in the control cell line KU812 or in the highly enriched eosinophil preparations. As expected, a 1.2-kilobase (kb) transcript is readily detected in KU812 cells (Fig. 3a). Eosinophils from all patients express transcripts of similar size, but with variable abundance according to the individual donor (Fig. 3a, lanes 2–5). No transcript is detected in preparations of lymphocytes or neutrophils (Fig. 3b, lanes 2–5), which may contaminate the eosinophil preparations. Furthermore, a similarly sized mRNA is also detected in butyrate-treated HL60 cells, which differentiate towards eosinophils^{16,17} (Fig. 3b, lanes 6 and 7), but not in untreated cells (Fig. 3a, lane 6; Fig. 3b, lane 8). This is remarkable not only because FcεRI α-chain transcripts are expressed in an eosinophil cell line, but also because this expression can be regulated.

FcεRI is a tetrameric structure composed of an IgE-binding α-chain, a β-chain and two disulphide-linked γ-chains^{8,9}. We tested for the presence of β- and γ-transcripts in our eosinophil preparations. As in basophils, the amount of β-chain mRNA is much less than for α-chain¹⁸, so the polymerase chain reaction

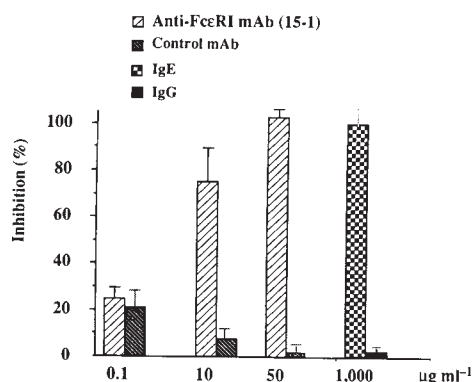


FIG. 1 Inhibition of ¹²⁵I-IgE binding to human eosinophils by anti-FcεRI α-chain monoclonal antibody (mAb) 15-1.

METHODS. Human eosinophils were purified from the venous blood of patients with eosinophilia, associated mainly with idiopathic hypereosinophilic syndrome (HES), skin diseases or lymphomas, by centrifugation on discontinuous 18–25% (w/v) metrizamide gradients (Nyegaard, Norway) as described²². The degree of purity of eosinophil populations, estimated after staining with Giemsa, was between 85 and 99%. Human IgE myeloma protein PS (gift from H. L. Spiegelberg) was radioiodinated by the chloramine-T technique⁷ to a specific activity of 1×10^6 c.p.m. per µg. Purified human eosinophils (3.3×10^6 cells per ml) were incubated with ¹²⁵I-labelled IgE myeloma protein ($1 \mu\text{g ml}^{-1}$) in a total volume of 60 µl for 90 min at 4 °C in the presence of various concentrations of anti-FcεRI α-chain mAb 15-1, or isotopic control mAb, or human IgE myeloma protein PS, or normal human IgG (Sigma). Cells were then centrifuged through sucrose gradients and cell-associated radiolabelled IgE was measured in a γ-scintillation counter. All results obtained with anti-FcεRI α-chain mAb 15-1 or control antibodies are presented as mean percentages of inhibition ± s.e.m. by comparison to those obtained with unlabelled IgE ($n=4$ experiments). In two experiments, the average numbers of receptors per eosinophil were estimated to be 84,000 and 73,000 respectively.

(PCR) was used to detect β -chain mRNA. The amplified products obtained from eosinophils, KU812 cells and, remarkably, from butyrate-treated HL60, revealed the size bands predicted for Fc ϵ RI β - and γ -chains (Fig. 3c, d). The specificity of the fragments was confirmed by restriction enzyme analysis (data not shown) and Southern blotting, with internal oligonucleotides specific for each subunit (Fig. 3e, f). Thus, eosinophils and

differentiated HL-60 cells express α -, β - and γ -transcripts.

Effector functions of eosinophils appear to be primarily mediated by the release of granule proteins¹⁹. Moreover, IgE-dependent activation of eosinophils from patients with parasitic infections or with allergic diseases can lead to the selective release of eosinophil peroxidase (EPO) but not ECP^{20,21}. We therefore investigated whether crosslinking of Fc ϵ RI could induce EPO

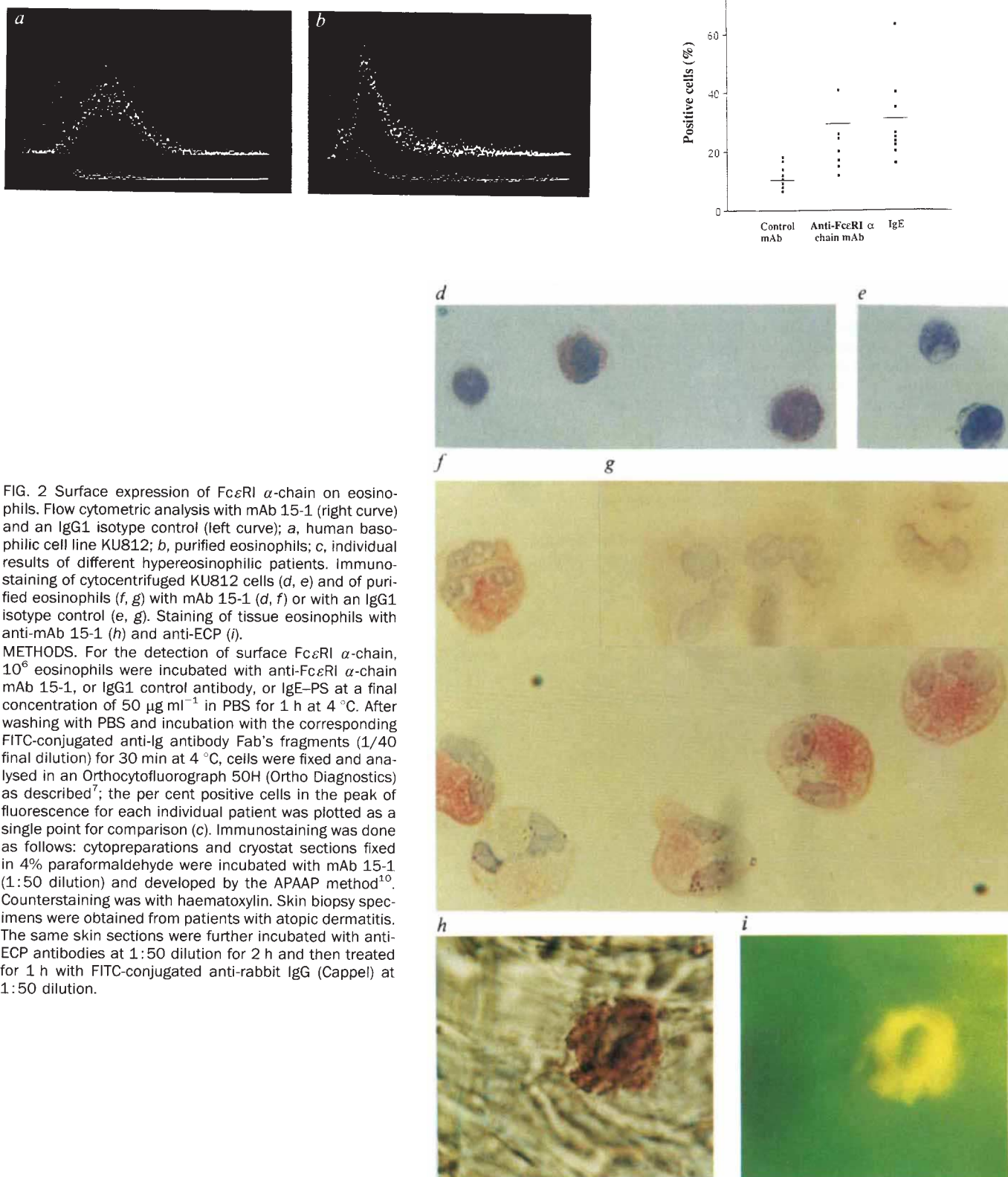
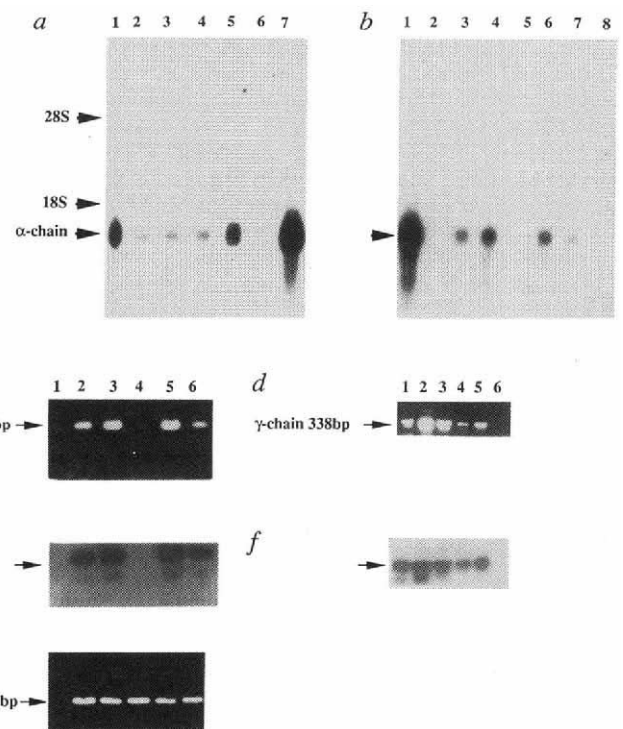


FIG. 2 Surface expression of Fc ϵ RI α -chain on eosinophils. Flow cytometric analysis with mAb 15-1 (right curve) and an IgG1 isotype control (left curve); a, human basophilic cell line KU812; b, purified eosinophils; c, individual results of different hypereosinophilic patients. Immunostaining of cytocentrifuged KU812 cells (d, e) and of purified eosinophils (f, g) with mAb 15-1 (d, f) or with an IgG1 isotype control (e, g). Staining of tissue eosinophils with anti-mAb 15-1 (h) and anti-ECP (i).

METHODS. For the detection of surface Fc ϵ RI α -chain, 10^6 eosinophils were incubated with anti-Fc ϵ RI α -chain mAb 15-1, or IgG1 control antibody, or IgE-PS at a final concentration of $50 \mu\text{g ml}^{-1}$ in PBS for 1 h at 4°C . After washing with PBS and incubation with the corresponding FITC-conjugated anti-Ig antibody Fab's fragments (1/40 final dilution) for 30 min at 4°C , cells were fixed and analysed in an Orthocytofluorograph 50H (Ortho Diagnostics) as described⁷; the per cent positive cells in the peak of fluorescence for each individual patient was plotted as a single point for comparison (c). Immunostaining was done as follows: cytopreparations and cryostat sections fixed in 4% paraformaldehyde were incubated with mAb 15-1 (1:50 dilution) and developed by the APAAP method¹⁰. Counterstaining was with haematoxylin. Skin biopsy specimens were obtained from patients with atopic dermatitis. The same skin sections were further incubated with anti-ECP antibodies at 1:50 dilution for 2 h and then treated for 1 h with FITC-conjugated anti-rabbit IgG (Cappel) at 1:50 dilution.

FIG. 3 Presence of Fc ϵ RI α -, β - and γ -chain transcripts in eosinophils. *a* and *b*, Northern blot hybridization of total RNA with a human Fc ϵ RI α -chain probe. *a*, Lanes: 1 and 7, human basophilic cell line KU812; 2–5, highly purified eosinophils from 4 different eosinophilic patients; 6, undifferentiated HL60 cells. *b*, Lanes: 1, KU812 cells; 2, purified lymphocytes from hyper-eosinophilic patient; 3 and 4, purified eosinophils, as in lanes 3 and 4 in *a*; 5, highly purified neutrophils from eosinophilic patient; 6 and 7, butyrate-induced HL60 cell line, at 7 and 3 days respectively; 8, undifferentiated HL60 cells. *c* and *e*, Reverse transcriptase (RT)-PCR and Southern blot analysis of Fc ϵ RI β -subunit. Lanes: 1, negative control (no cDNA); 2 and 6, purified eosinophils; 3, KU812; 4, undifferentiated HL60 cells; 5, differentiated HL60 cells. *d* and *f*, RT-PCR and Southern blot analysis of Fc ϵ RI γ -subunit. Lanes: 1 and 4, purified eosinophils; 2, differentiated HL60 cells; 3, undifferentiated HL60 cells; 5, KU812 cells; 6, negative control (no cDNA). *g*, RT-PCR with β -actin primers.

METHODS. Neutrophil preparations: patient neutrophils were obtained from the venous blood of patients with hyper-eosinophilia by centrifugation through discontinuous metrizamide gradients²³. Cell culture: a sub-line of the HL60 cell line (clone 15; gift from J. Tavernier (Roche Research, Gent, Belgium) was maintained in RPMI 1640 medium supplemented with 10% FCS and 25 $\mu\text{g ml}^{-1}$ gentamicin. For induction, cells were incubated at 2×10^5 per ml with 0.5 mM sodium butyrate for 7 days¹⁷. For northern blot analysis, 10 μg total RNA was used per lane⁹. The *Bam*HI-*Xho*I fragment of the human Fc ϵ RI α -chain cDNA was used as the probe. RT-PCR: total RNAs from eosinophils, KU812 and HL60 cell line were reverse-transcribed and examined by PCR analysis using specific primers for β - and γ -chains of Fc ϵ RI (ref. 11). PCR conditions were 1 min at 94 °C, 2 min at 60 °C and 3 min at 72 °C for 30 cycles. PCR products were electrophoresed on agarose and visualized by ethidium staining. Southern blot analysis: Amplified products were blotted onto Hybond N membranes using standard methods. Oligonucleotide probes were labelled with T4 polynucleotide kinase and [γ -³²P]ATP. Hybridization was carried out in the same conditions as before.



Oligonucleotides specific for sequences on either side of a splice junction within β and γ genes were used in the PCR reaction to preclude amplification of possible contaminating genomic DNA. The Fc ϵ RI β - and γ -chain primers have been described¹¹.

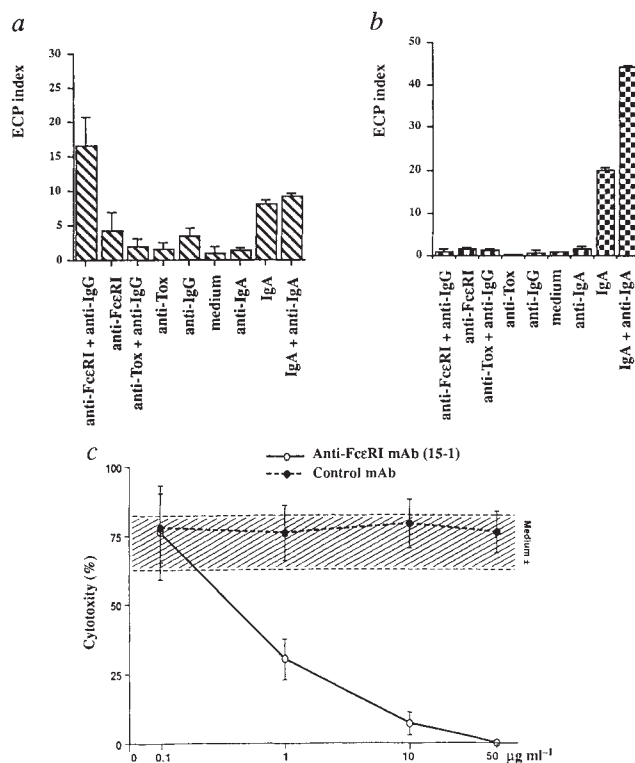
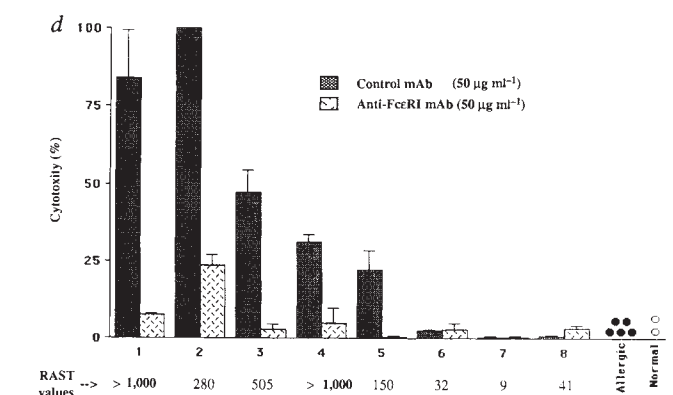


FIG. 4 *a* and *b*, Release of EPO and ECP after stimulation of eosinophils with anti-Fc ϵ RI mAb. *c* and *d*, Inhibition of eosinophil-mediated IgE-dependent cytotoxicity against schistosome targets. For cell activation, highly purified eosinophils ($>90\%$) were incubated with anti-Fc ϵ RI α -chain mAb 15-1, irrelevant mAb of the same isotype, or secretory IgA (positive control) at a final concentration of 10 $\mu\text{g ml}^{-1}$.



at 4 °C. Then cells were challenged with anti-Ig antibody Fab/2 fragments and incubated for 2 h at 37 °C in 5% CO₂. As negative controls, cells were incubated with each reagent alone. EPO and ECP release was evaluated in the same supernatants by a chemiluminescence assay²⁰ and a double-antibody radioimmunoassay method (Pharmacia Diagnostics) respectively. Results are expressed as the index of EPO and ECP release as described²¹. For cytotoxicity assay, eosinophils were preincubated with various concentrations of anti-Fc ϵ RI mAb 15-1 or isotype control for 30 min at room temperature before addition to *S. mansoni* targets³, in the presence of one IgE containing reference immune serum (at 1:32 final dilution; total volume, 200 μl). Results were compared with those obtained with targets incubated without mAb (medium) in the presence of effector cells and immune serum. The cytotoxicity values obtained with eosinophils from 4 different eosinophilic donors were expressed as mean % cytotoxicity $\pm$ s.e.m. *d*, Individual results of sera (final dilution 1:32) from 8 different *S. mansoni*-infected patients (patients 1–8), containing varying amounts of specific anti-*S. mansoni* IgE antibodies, were evaluated by a radioallergosorbent test against a soluble *S. mansoni* worm extract. IgE values are expressed as the index of bound radioactivity versus normal serum (RAST values). Anti-Fc ϵ RI (15-1) or control isotype mAb were used as in *c* (mean cytotoxicity values of 3 experiments $\pm$ s.e.m.). Sera from 5 allergic individuals containing between 420 and 4,313 IU ml⁻¹ total IgE were used as specificity controls. Normal human sera were used as negative controls.

release from eosinophils (Fig. 4a). Increased levels of EPO (index of release >15) are detected after stimulation of purified eosinophils by crosslinking antibody 15-1 but not an unrelated isotype control antibody, or other control antibodies. In contrast, no significant amounts of ECP were detected in the same supernatants in which EPO was measured (Fig. 4b). However, elevated levels of both EPO and ECP are obtained in supernatants of the same eosinophil preparations after crosslinking of secretory IgA (Fig. 4a, b). These data indicate that FcεRI is involved in the differential release of EPO versus ECP, as was shown previously after IgE-dependent activation^{20,21}.

Finally, we analysed whether FcεRI is involved in the cytotoxicity of eosinophils against parasites. Purified eosinophils, schistosomula targets, and IgE antibody-containing immune sera were incubated in the presence of antibody 15-1 or an isotype control antibody. A dose-dependent inhibition of cytotoxicity is

observed with 15-1 but not with control antibody in the presence of one immune serum used as positive control (Fig. 4c). Also, when more sera from *S. mansoni*-infected patients were used, not only were cytotoxicity levels significantly correlated with specific anti-*S. mansoni* IgE levels (Spearman's correlation test, $P < 0.05$), but they were also significantly inhibited by 15-1 at the optimum concentration (Fig. 4d). None of the five control isotype sera from allergic patients was able to induce target cell lysis. Anti-CD23 monoclonal antibody of the same isotype (antibody 135)⁷ used at $50 \mu\text{g ml}^{-1}$ did not significantly inhibit cytotoxicity (data not shown). These results indicate that FcεRI can mediate eosinophil-dependent cytotoxicity against schistosomula and provide the evidence that, in addition to its role in mediating allergic responses^{8,7}, FcεRI may participate in a physiological protective immune response against metazoan parasites. □

Received 29 July; accepted 16 November 1993.

1. Capron, A. et al. *Immun. Today* **7**, 15–18 (1986).
2. Capron, A., Dessaint, J. P., Capron, M., Ouma, J. H. & Butterworth, A. E. *Science* **238**, 1065–1072 (1987).
3. Hagan, P., Blumenthal, U. J., Dunne, D., Simpson, A. J. & Wilkins, H. A. *Nature* **349**, 243–245 (1991).
4. Rihet, P., Demeure, C. E., Bourgeois, A., Prata, A. & Dessein, A. J. *Eur. J. Immun.* **21**, 2679–2686 (1991).
5. Dunne, D. W. et al. *Eur. J. Immun.* **22**, 1483–1486 (1992).
6. Capron, M., Grangette, C., Torpier, G. & Capron, A. *Chem. Immun.* **47**, 128–178 (1989).
7. Capron, M. et al. *Eur. J. Immun.* **21**, 2423–2429 (1991).
8. Kinet, J. P. in *Inflammation: Basic Principles and Clinical Correlates* 701–708 (Raven, New York, 1992).
9. Ravetch, J. V. & Kinet, J. P. *Rev. Immun.* **9**, 457–479 (1991).
10. Wang, B. et al. *J. exp. Med.* **175**, 1353–1358 (1992).
11. Bieber, T. et al. *J. exp. Med.* **175**, 1285–1291 (1992).
12. Kochan, J., Pettine, L. F., Hakimi, J., Kisch, K. & Kinet, J. P. *Nucleic Acids Res.* **16**, 3584–3594 (1988).

13. Shimizu, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1907–1911 (1988).
14. Blank, U., Ra, C. & Kinet, J. P. *J. biol. Chem.* **266**, 2639–2646 (1991).
15. Hakimi, J. et al. *J. biol. Chem.* **265**, 22079–22081 (1990).
16. Fischkoff, S. A. et al. *J. exp. Med.* **160**, 179–196 (1984).
17. Plaetinck, G. et al. *J. exp. Med.* **172**, 683–691 (1990).
18. Kuster, H., Zhang, L., Brini, A. T., MacGlashan, D. W. J. & Kinet, J. P. *J. biol. Chem.* **267**, 12782–12787 (1992).
19. Gleich, G. J. & Adolphson, C. R. *Adv. Immun.* **39**, 177–253 (1986).
20. Khalife, J. et al. *J. Immun.* **137**, 1659–1664 (1986).
21. Tomassini, M. et al. *J. Allerg. clin. Immun.* **88**, 365–375 (1991).
22. Capron, M. et al. *J. exp. Med.* **164**, 72–89 (1986).
23. Truong, M. J. et al. *J. exp. Med.* **177**, 243–248 (1993).

ACKNOWLEDGEMENTS. We thank J. P. Dessaint, M. Joseph and M. H. Jouvin for advice; A. Scharenberg for critically reading the manuscript; J. P. Papin, S. Loiseau, J. L. Neyrink and J. P. Kusnier for help with experimental procedures; K. Nilsson for the KU812 cell line; and J. Pestel and P. Desreumaux for human sera. A. S. G. is supported by Fondation Marcel Mérieux, Lyon, France.

Diadenosine phosphates and the physiological control of blood pressure

Hartmut Schlüter, Elmar Offers, Gitta Brüggemann, Marcus van der Giet, Martin Tepel, Eckhard Nordhoff*, Michael Karas*, Claus Spieker, Herbert Witzel† & Walter Zidek

Medizinische Poliklinik, *Institut für Medizinische Physik, † Institut für Biochemie der Universität Münster, Albert Schweitzer Strasse 33, Münster 48129, Germany

OUR understanding of the regulation of vascular tone has been extended since the identification of vasoactive agents such as the atrial natriuretic peptides¹, endothelial-derived relaxing factor² and endothelin³. Unidentified vasopressors have been found in platelets⁴. Here we isolate these vasopressors and identify them as diadenosine pentaphosphate (AP₅A) and diadenosine hexaphosphate (AP₆A) by chromatography, mass spectrometry, ultraviolet spectroscopy and enzymatic cleavage. In the vasculature of isolated perfused rat kidney, both diadenosine phosphates were active at a concentration of 10^{-9} M; in aortic rings, contractions were elicited at 10^{-8} M. Intra-aortic injection in the rat caused a prolonged increase in blood pressure. We conclude that AP₅A and AP₆A may play a part in local vasoregulation and possibly in the regulation of blood pressure.

The vasopressors were purified from platelet concentrates in the following steps (Fig. 1): (1) deproteinization with perchloric acid; (2) binding to a preparative C18 column using triethylammonium acetate as a cationic ion-pair reagent; (3) displacement chromatography on a second C18 column (Fig. 1a); (4)

displacement chromatography on an anion-exchange column (Fig. 1b); and (5) gradient elution with triethylammonium acetate as a cationic ion-pair reagent on a third C18 column (Fig. 1c, d). After each chromatographic step, vasopressor activity of the fractions was assayed using an isolated perfused rat kidney (Fig. 1 and Fig. 2a, b, g). After the last purification step, a single homogeneous peak was obtained (Fig. 1e, f).

The active substances were identified as diadenosine pentaphosphate (AP₅A) and diadenosine hexaphosphate (AP₆A) by the following findings (Fig. 3): (1) matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) revealed relative molecular masses of 916 (Fig. 3a) and 996 (Fig. 3b); (2) after addition of Na⁺, MALDI mass spectrometry showed that there were five and six negative charges in AP₅A and AP₆A, respectively (Fig. 3c, d); (3) ultraviolet spectroscopy gave a spectrum characteristic of adenine (Fig. 3e, f); (4) the retention times of the active fractions in step 5 were identical to those of commercially available AP₅A and AP₆A; and (5) MALDI-MS of commercially available AP₅A and AP₆A with and without a Na⁺ salt gave the same pattern as the fractions in step 5. The structures of AP₅A and AP₆A were evidenced by the following results (Fig. 3c, d, g and h): (1) MALDI analysis revealed five and six negative charges, respectively; however, isomers of AP₅A and AP₆A with one terminal phosphate group in the 3' position would have six and seven negative charges as one terminal phosphate group would introduce two negative charges into the molecule; (2) both compounds were resistant to alkaline phosphatase; (3) snake venom phosphodiesterase (5' nucleotidase) yielded AMP plus AP₄ and AP₅, respectively, whereas calf spleen phosphodiesterase (3' nucleotidase), which is specific for 3' linkages, was ineffective. The molar ratio of ATP to AP₅A and AP₆A in the platelet extracts was estimated to be in the range of 50:1 and 100:1. ATP content was estimated chromatographically⁵. When platelets were aggregated with 1 U ml^{-1} thrombin (Sigma), AP₅A and AP₆A were isolated from the supernatant by the purification procedure already described.

The following effects of AP₅A and AP₆A on vascular tone were observed (Fig. 2): (1) vasoconstriction in the vasculature of the isolated perfused rat kidney at concentrations of $\geq 10^{-9}$ M; (2) isometric contraction of aortic rings at $\geq 10^{-8}$ M; (3) increase in blood pressure after intra-aortic injection of 10^{-7} mol in the rat; and (4) increase in cytosolic free Ca²⁺ concentration in rat aorta. A bioassay of the active fractions from step 5 of the purification procedure (Fig. 1d) is shown for comparison (Fig. 2g).

AP₅A and AP₆A have been found respectively in bovine adrenal tissue^{6,7} and chromaffin cells⁸, although neither has so far been reported in human tissue. These diadenosine polyphosphates can act as competitive inhibitors of ADP-induced platelet aggregation and of the platelet release reaction⁹. AP₄A, AP₅A

and AP₆A stimulate Ca²⁺ release from skeletal muscle sarcoplasmic reticulum¹⁰, their potency increasing with the number of phosphate groups¹⁰.

Although little is known about AP₅A and AP₆A, other compounds of this type, namely AP₃A and AP₄A, have been investigated for their effects on vasculature. Both agents dilate coronary and mesenteric arteries¹¹⁻¹³, and in addition to these extracellular activities, AP₄A may act as an intracellular growth-promoting signal^{14,16}. Although aminoacyl transfer RNA synthetase is involved in the formation of AP₃A and AP₄A (*in vitro*)¹⁷, the biosynthetic pathways are unknown for AP₅A and AP₆A. Platelets from essential hypertensives have higher vasopressor activity in the fractions containing AP₅A and AP₆A (ref. 4) and both AP₅A and AP₆A are released during platelet aggre-

FIG. 1 Purification of the diadenosine phosphates. **a**, Step 3: reversed-phase HPLC (LiChrosorb RP-18, 250 × 4 mm; Merck, Darmstadt, Germany) in the displacement mode¹⁸ (equilibration and sample buffer: 20 mM triethylammonium acetate (TEAA) in water (eluent 1); displacer eluent: eluent 1 plus 160 mM butanol; flow rate 100 μ l min⁻¹). Bar indicates region of vasopressor activity. **b**, Step 4: anion-exchange displacement¹⁸ chromatography (MonoQ PC1.6/5, chloride form; Pharmacia, Freiburg, Germany) of the vasopressor fraction from **a** dissolved in water at pH 7; sample volume was 1 ml (displacer eluent: 30 mM citrate, pH 7, 100% after 100 min; flow rate: 10 μ l min⁻¹). $\uparrow$, Vasopressor activity. **c** and **d**, Step 5: reversed-phase gradient chromatography (Superspher RP-18 endcapped, 250 × 4 mm (Merck); eluent 1: as in **a**; eluent 2: 20 mM TEAA in acetonitrile; the gradient is shown as a dotted line) of the vasopressor fractions from **b** (**c**, first arrow and **d**, second arrow in **b**). Fractions tested in the bioassay shown in Fig. 2g are indicated by vertical dotted lines. **e** and **f**, Rechromatography of the vasoactive fractions from **c** and **d** (reversed-phase gradient chromatography using Lichrospher RP-select B (250 × 4 mm) from Merck; eluent 1: tetrabutylammoniumhydrogensulphate (TBA; 2 mM), K₂HPO₄ in water (10 mM); eluent 2: TBA (2 mM) in 80% acetonitrile in water; gradient: 0 min, 0% eluent 2; 2 min, 31% eluent 2; 25 min, 36% eluent 2; 27 min, 80% eluent 2; 30 min, 80% eluent 2; 32 min, 0% eluent 2). Retention times: AP₅A, 27.5 min; AP₆A, 28.2 min.

METHODS. The supernatant from platelets (from 4.5 l blood) deproteinized with perchloric acid⁴ (step 1) was mixed with TEAA (final concentration 20 mM) and loaded onto a reversed-phase column (step 2; LiChroprep RP-18 A from Merck; equilibration and sample buffer: eluent 1; elution buffer: 50 ml eluent 2). Vasopressor activity of the fractions was assayed with an isolated rat kidney perfused with constant flow⁴.

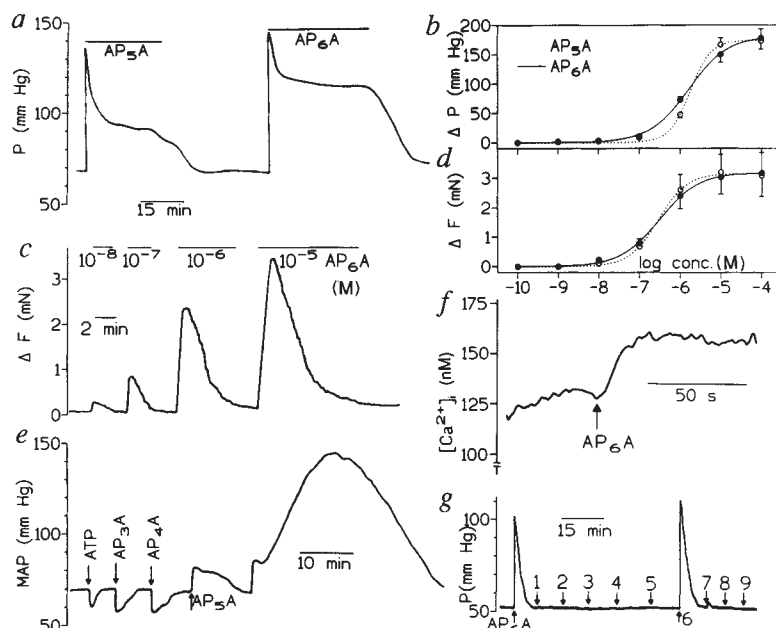
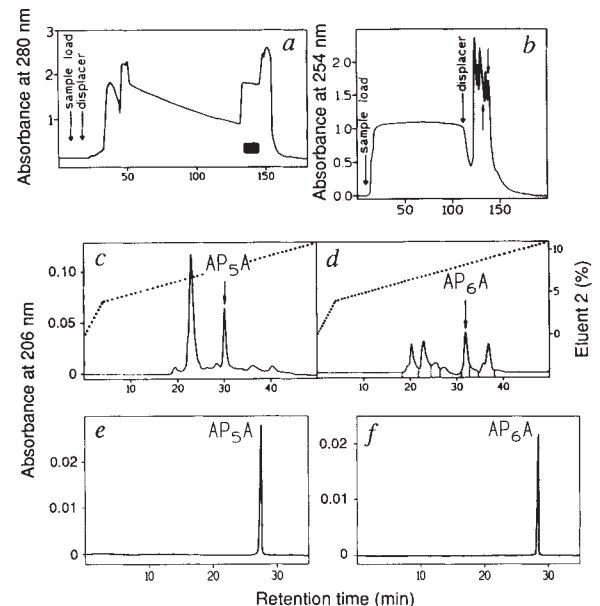


FIG. 2 Vasopressor actions of AP₅A and AP₆A (Sigma). **a**, Perfusion pressure (*P*) of an isolated rat kidney⁴ perfused with constant flow during perfusion with AP₅A and AP₆A (horizontal line). **b**, Concentration-response curve of the experiments shown in **a** for both AP₅A and AP₆A (each *n* = 7). **c**, Isometric contraction of an aortic ring¹⁹ by AP₆A (ΔF , force); incubation periods indicated by horizontal lines. **d**, Concentration-response curve (mean $\pm$ s.d.) of the maximum contractions as shown in **c** for both AP₅A and AP₆A (each *n* = 9). **e**, Mean arterial pressure (MAP) in a rat¹⁹ after intra-aortic injection (arrows) of 5×10^{-6} mol ATP, 10^{-7} mol AP₃A, 10^{-7} mol AP₄A, 10^{-7} mol AP₅A and 10^{-7} mol AP₆A (each dissolved in 100 μ l physiological salt solution). **f**, Changes in cytosolic free Ca²⁺ concentration ($[Ca^{2+}]_i$) induced by 10^{-6} M AP₆A measured with Fura 2 (refs 20, 21) in a segment of rat aorta. **g**, Perfusion pressure (*P*) of an isolated rat kidney⁴ perfused with constant flow during bolus injections of AP₆A (0.5 nmol) and of 9 chromatographic fractions (each dissolved in 100 μ l physiological salt solution) from step 5 of the purification procedure (Fig. 1d).

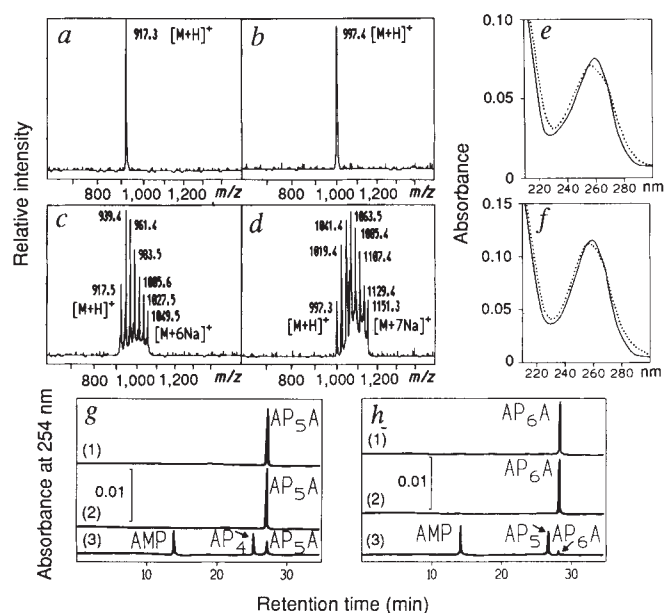


FIG. 3 Identification of the vasopressor agents. *a-d*, Positive-ion mass spectra measured with the matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)²² of the vasopressor fractions shown in Fig. 1c and d (abscissa shows relative mass/charge, m/z). MALDI-MS produces mainly singly charged ion species. This is either achieved by protonation of the analyte molecules, when the signals reveal the molecular mass plus 1, or in salt-containing media, by attachment of metal cations (Na^+ in c and d)²². In *a* and *b*, $m/z=917$ and $m/z=997$ correspond to the molecular masses of the protonated ions of AP_5A and of AP_6A ; c and d show MALDI spectra of the fractions in *a* and *b* with 10 mM NaCl in the analyte solution. According to the number of negative charges, 6 and 7 Na^+ ions at most can be attached to AP_5A and AP_6A to produce singly charged molecules, generating 6 and 7 additional signals. *e* and *f*, Ultraviolet spectra of the vasopressor fraction shown in Fig. 1c and d: solid line, in water at pH 6.4, maximum at 260 nm; dotted line, at pH 2 in 0.01 M HCl, maximum at 257 nm. Spectra correspond with those of adenine under the same conditions. *g* and *h*, Reversed-phase chromatography (conditions as described for Fig. 1e) after incubation of the vasoactive fractions from the reversed phase shown in Fig. 1c and d with different enzymes at 37 °C (Boehringer Mannheim): (1) 1 mU ml^{-1} alkaline phosphatase (EC 3.1.3.1; incubation for 1 h); (2) 1 mU ml^{-1} phosphodiesterase (3' nucleotidase, EC 3.1.16.1; incubation, 1 h); and (3) 3 mU ml^{-1} phosphodiesterase (5' nucleotidase, EC 3.1.15.1, purified according to ref. 23; incubation, 9 min). Absorbance range is indicated by a bracket. Retention times: AMP, 13.9 min; AP_4 , 25.2 min; AP_5 , 26.5 min; AP_5A , 27.4 min; AP_6A , 28.0 min. Peaks were identified by comparing retention times with those of authentic standards (except AP_5) and by MALDI-MS.

gation, so the role of AP_5A and AP_6A in primary hypertension will have to be studied in more detail. Furthermore, neither AP_5A nor AP_6A has been tested for growth stimulatory properties, which apart from their vasopressive activity could also play a role in the induction of vascular hypertrophy, an obligatory companion of hypertension. □

Received 2 August; accepted 22 October 1993.

1. Flynn, T. G., De Bold, M. L. & De Bold, A. J. *Biochem. biophys. Res. Commun.* **117**, 859–865 (1983).
2. Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524–526 (1987).
3. Yanagisawa, M. et al. *Nature* **332**, 411–415 (1988).
4. Agha, A. et al. *J. Vasc. Res.* **29**, 281–289 (1992).
5. Ally, A. & Park, G. J. *Chromatogr.* **575**, 19–27 (1992).
6. Pintor, J., Torres, M. & Miras-Portugal, M. T. *Life Sci.* **48**, 2317–2324 (1991).
7. Rodríguez del Castillo, A., Torres, M., Delicado, E. G. & Miras-Portugal, M. T. *J. Neurochem.* **51**, 1696–1703 (1988).
8. Pintor, J., Rotllán, P., Torres, M. & Miras-Portugal, M. T. *Analyt. Biochem.* **200**, 296–300 (1992).

9. Harrison, M. J. & Brossmer, R. *FEBS Lett.* **54**, 57–60 (1975).
10. Morii, H. & Makino, M. *Eur. J. Biochem.* **205**, 979–984 (1992).
11. Pohl, U., Ogilvie, A., Lamontagne, D. & Busse, R. *Am. J. Physiol.* **260**, H1692–H1697 (1991).
12. Nees, S. *Eur. Heart J.* **10**, suppl. F 28–35 (1989).
13. Busse, R., Ogilvie, A. & Pohl, U. *Am. J. Physiol.* **254**, H828–H832 (1988).
14. Baril, E. F., Coughlin, S. A. & Zamecnik, P. C. *Cancer Invest.* **3**, 465–471 (1985).
15. Yamaguchi, N., Kodama, M. & Ueda, K. *Gastroenterology* **89**, 723–731 (1985).
16. Weinmann-Dorsch, C. et al. *Eur. J. Biochem.* **138**, 179–185 (1984).
17. Zamecnik, P. C., Stephenson, M. L., Janeway, C. M. & Randerath, K. *Biochem. biophys. Res. Commun.* **24**, 91–97 (1966).
18. Frenz, J., Van der Schrieck, P. & Horvath, C. J. *Chromatogr.* **330**, 1–17 (1985).
19. Zidek, W. et al. *Clin. Exp. Hypertens.* **A12**, 365–381 (1990).
20. Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440–3450 (1985).
21. Neusser, M., Golinski, P., Zhu, Z., Tepel, M. & Zidek, W. *J. Vasc. Res.* **30**, 116–120 (1993).
22. Nordhoff, E. et al. *Rapid Comm. Mass Spectrom.* **6**, 771–776 (1992).
23. Sulikowski, E. & Laskowski, M. *Biochim. biophys. Acta* **240**, 443–447 (1971).

ACKNOWLEDGEMENTS. This work was supported by the DFG. We thank W. Sibrowski for platelet concentrates.

Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice

Stephanie M. Toggas*, Eliezer Masliah†, Edward M. Rockenstein*, Glenn F. Rall*, Carmela R. Abraham‡ & Lennart Mucke*§

* Department of Neuropharmacology, Division of Virology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

† Departments of Neurosciences and Pathology, University of California at San Diego, La Jolla, California 92093-0624, USA

‡ Departments of Medicine and Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118-2394, USA

MANY people infected with human immunodeficiency virus type 1 (HIV-1) develop neurological complications that can culminate in dementia and paralysis¹. The discrepancy between the severity of impairment and the paucity of detectable HIV-1 within neurons has led to an intense search for diffusible virus- and host-derived factors that might be neurotoxic (see ref. 2 for review). The HIV-1 envelope glycoprotein gp120 is an extracellular protein that is shed from infected cells³ and so has the potential to diffuse and interact with distant uninfected brain cells. Studies on cultured immature cells suggest that gp120 induces neurotoxicity (reviewed in refs 2, 4), and systemic injection of gp120 in neonatal rats⁵ and intracerebroventricular injection in adult rats results in deleterious effects on the brain^{6,7}. To assess the pathogenic potential of gp120 in the intact brain, we have now produced gp120 in the brains of transgenic mice and found a spectrum of neuronal and glial changes resembling abnormalities in brains of HIV-1-infected humans. The severity of damage correlated positively with the brain level of gp120 expression. These results provide *in vivo* evidence that gp120 plays a key part in HIV-1-associated nervous system impairment. This model should facilitate the evaluation and development of therapeutic strategies aimed at HIV–brain interactions.

The gp120 protein was expressed in astrocytes of transgenic mice. The *env* sequence encoding gp120 was placed under the regulatory control of a modified murine glial fibrillary acidic protein (GFAP) gene (Fig. 1a). Because GFAP is expressed predominantly in astrocytes, cells that comprise a substantial proportion of the central nervous system (CNS), this strategy provides for effective transgene expression within the CNS^{8–10}.

§ To whom correspondence should be addressed.

Culture media from astroglial cells stably transfected with the GFAP-gp120 construct contained gp120 (Fig. 1b), confirming production of secreted gp120 from this transgene.

Eleven GFAP-gp120 transgenic founder mice were obtained from microinjections of fertilized murine (B6×SJL) oocytes;

five founders were used to establish separate transgenic lines (lines 2, 3, 10, 16 and 37). Lines 2, 10, and 16 were analysed in detail. CNS levels of gp120 messenger RNA were similar in mice from the same line but differed across lines (Fig. 1c) (2–11 mice per line). There was no correlation between levels of expression

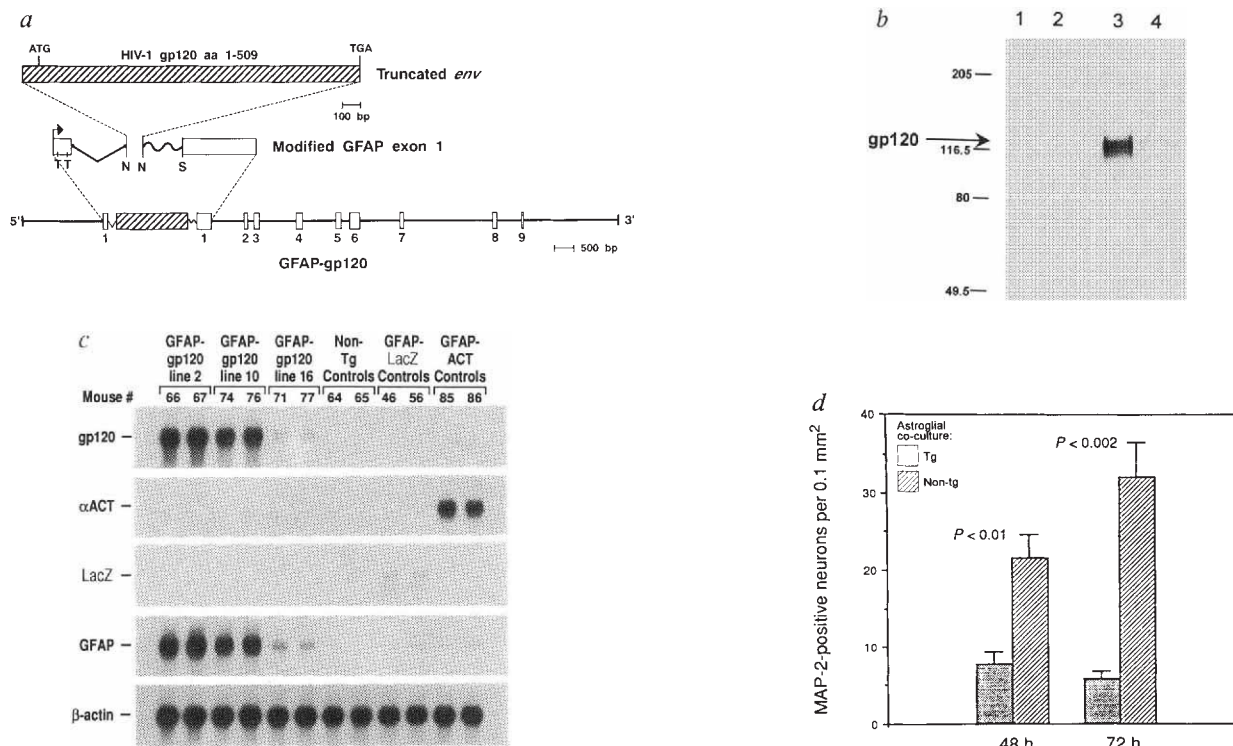
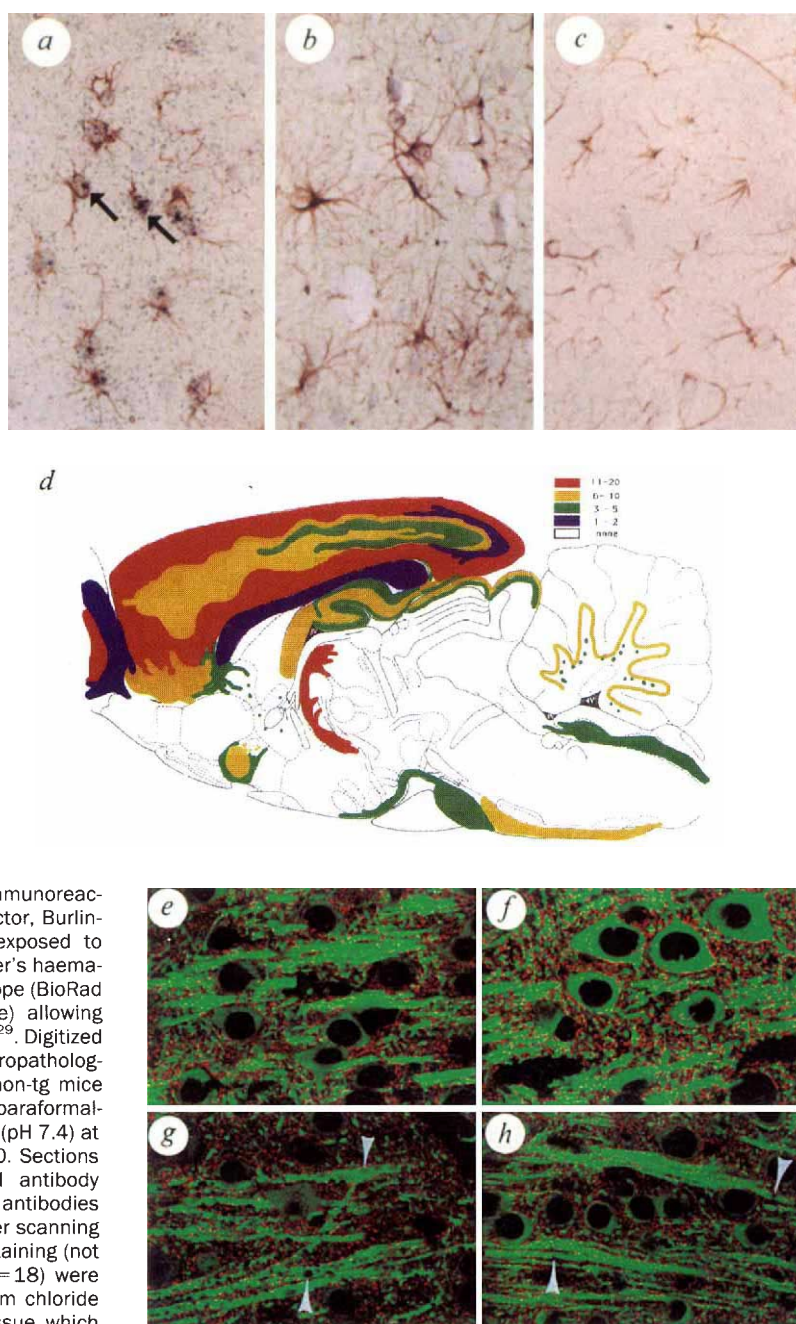


FIG. 1 a, GFAP-gp120 fusion gene. The *env* gene of HIV-1_{LAV} (ref. 22), truncated to encode gp120 amino acids 1–509, was inserted into exon 1 of a murine GFAP gene modified as follows. The ATGs between the transcriptional start site (arrow) and the unique *Sall* site (S) in GFAP exon 1²³ were mutated to TTGs (T). A Simian virus (SV40) late gene splice (V) and polyadenylation signal (A) were inserted 5' and 3' of a unique *NotI* site (N), respectively. Numbered boxes, GFAP exons. **b**, After metabolic labelling, immunoprecipitable gp120 was found in culture media from high expressor C6 astrocytoma cells (lane 3) stably transfected with the construct in **a** and from *env*-expressing HL2/3 cells (lane 4); no gp120 was seen in media from mock-transfected C6 (lane 1) or untransfected HeLa (lane 2) cells. Transfected C6 cells were used to demonstrate the functionality of the transgene because immunoprecipitation was not sensitive enough to detect gp120 in primary cultures of GFAP-gp120 transgenic astrocytes, whose gp120 mRNA levels were substantially lower than those found in transfected C6 and HL2/3 cells (data not shown). Numbers on the left represent relative molecular masses ($M_r \times 10^{-3}$). **c**, Northern analysis of poly (A)⁺-enriched brain RNA from 6-month-old mice. Groups were: GFAP-gp120 lines 2, 10 and 16, non-transgenic (Non-tg) littermates and transgenic (tg) mice expressing *Escherichia coli* lacZ (GFAP-lacZ line 445–7) or a human α 1-antichymotrypsin (ACT) complementary DNA (GFAP-ACT line 528–13)⁹ from the same GFAP vector as in **a**. **d**, Exposure to tg astrocyte cultures induced premature death in non-tg hippocampal neurons. Neurons were co-cultured with astrocytes from tg mice (solid bars) or from non-tg controls (hatched bars). The average number of MAP-2-positive cells remaining after 48 or 72 h of coculture was expressed as neurons per 0.1 mm². **METHODS**. Standard protocols were used for routine molecular biological techniques. Two overlapping regions of the murine GFAP genomic clone pSVGFA (obtained from N. J. Cowan, New York University Medical Center) were amplified in separate polymerase chain reactions (PCRs) using primers A + B and C + D, respectively. Primer sequences (5' to 3'): A = ATTCTGCAGCTAGAACTTGGGTGGGTCTTC; B = GGTGATGCGTCTCCGCTCCAACCTGCCCTG; C = ACGCATCACCTCTGCGCGCGCTCCTTTGC; D = ACGTGGCATGGTACCCAGTTGTGCACTAGG; E = AATGAGTCCGAGA TCTTCAGA; and F = GGATCCGGATCCGTCGACTCATTCTCTGCACCACT

CTTCT. Primers B and C introduced the two A to T mutations. The two amplicons were mixed and amplified with outside primers A + D. The resulting amplicon was subcloned into pCR1000 (In Vitrogen, La Jolla, California) for sequence verification, isolated by *PstI/Sall* digests and substituted for the corresponding region of the wild-type GFAP gene. The *XhoI/Sall* segment (containing the SV40 late gene splice and polyadenylation signal) of pCMV β was ligated into the GFAP *Sall* site and the lacZ element of pCMV β deleted by *NotI* digest. PCR amplification of HIV-1 *env* (*HindIII/HindIII* segment of clone pHenv²²) using primers E + F introduced a TGA stop codon following the Glu residue within the predominant endoproteolytic cleavage site of gp160 (Arg-Glu-Lys-Arg)²⁴. The 162-base pair (bp) amplicon was substituted for the original *env* 520-bp *BglIII/HindIII* segment. The truncated *env* gene was incorporated into the GFAP construct using *NotI* linkers, and the transgene purified and injected into fertilized B6×SJL oocytes as described⁸. Metabolic labelling and immunoprecipitations were performed essentially as described²⁵ using Expre³⁵S³⁵S Protein Labelling mix (NEN), polyclonal antisera to gp160 or gp120 (1:250 and 1:125 dilutions) and Gamma-Bind G Sepharose (Pharmacia). Signals were visualized using a Phosphorimager SF (Molecular Dynamics, Sunnyvale, California). RNA extraction and northern analysis were done as described⁸. The following DNAs were ³²P-labelled with a Prime-It II random primer kit (Stratagene, La Jolla, California): HIV-1 *env*, bp 5,608–7,328 (ref. 26); 2.7-kilobase (kb) rat GFAP cDNA clone rGFA15 (from R. Milner, Hershey Medical Center, Hershey, Pennsylvania), a 1.6-kb human ACT cDNA, *E. coli* lacZ, bp 105–2,661 (ref. 8), and mouse β -actin, bp 480–740 (ref. 8). Primary cultures of non-tg fetal hippocampal neurons (gestational day 18–19) were obtained as described²⁷. Single cell suspensions were plated at low density onto poly-L-lysine-coated Nunc inserts (polycarbonate, 0.2 μ m) and transferred onto confluent astrocytes in serum-free Bottenstein's N2 formulation (Gibco/BRL) for 48–72 h. Astrocyte cultures, obtained as described²⁰, contained 95% GFAP(+) astrocytes and 5% F4/80(+) microglia. Neuronal cultures were immunostained for MAP-2, scanned and quantified (10 fields analyzed per insert) as in Figs 2 and 4.

FIG. 2 Astroglial expression of gp120 mRNA and associated neuronal damage in GFAP-gp120 transgenic mice. **a–d**, Brain sections were co-labelled with gp120 anti-sense (**a**, **c**, **d**) or sense (**b**) riboprobes and antibodies against GFAP. **a**, gp120 mRNA signal (clusters of silver grains) over cells (arrows) stained for the astroglial marker GFAP (brown) (hippocampus). No astrocytes were radio-labelled with the corresponding sense probe (**b**) or with the antisense probe in non-transgenic (non-tg) sections (**c**). The distribution of radiolabelled astrocytes was mapped by computer-assisted image analysis (**d**); colours represent numbers of radiolabelled astrocytes per 0.1 mm^2 . **e–h**, Laser confocal images of neocortical sections show neuronal cell bodies and dendrites (green) and presynaptic terminal boutons (red). In non-tg (**e**) and GFAP-lacZ transgenic (tg) (**f**) controls, antibodies against MAP-2 (green) labelled normal pyramidal neurons and apical dendrites. Antibodies against synaptophysin (red) labelled presynaptic terminals in a typical dense, punctate pattern. In contrast, dilation and severe vacuolization of dendrites (arrows) and decreases in synapto-dendritic complexity were seen in the high expressor tg line 2 (**g**). A less severe form of this damage was found in the low expressor tg line 16 (**h**).

METHODS. Six-month-old tg mice of line 2 ($n=3$) and non-tg controls ($n=3$) were anaesthetized with chloral hydrate and perfused transcardially with ice-cold phosphate-buffered 4% paraformaldehyde. Brains were postfixed in 10% buffered zinc formalin, paraffin-embedded and processed for *in situ* hybridization as described²⁸. Sequences encoding gp120 amino acids 1–42 and 466–509 were fused and subcloned into pSP70 for synthesis of ³⁵S-labelled RNA probes using the Riboprobe System (Promega). Sections ($3 \mu\text{m}$) from tg and non-tg mice were mounted on the same slide and hybridized at $50\text{--}55^\circ\text{C}$ essentially as described²⁸. GFAP antibodies (Dako, Carpinteria, CA) were used at a 1:1,000 dilution in 0.05 M Tris (pH 7.6)/0.26 M NaCl/1.5% fetal bovine serum, and immunoreactivity was detected with an avidin–biotin peroxidase kit (Vector, Burlingame, CA) and diaminobenzidine/ H_2O_2 . Sections were exposed to Kodak NTB-2 emulsion for 7 days, counterstained with Mayer's haematoxylin and scanned with a laser scanning confocal microscope (BioRad MRC-600) using an optical configuration (reflection cube) allowing detection of silver grains associated with GFAP-positive cells²⁹. Digitized images were quantified using Image 1.23 software. For neuropathological analyses, brains from 1.5- to 2.5-month-old tg and non-tg mice were flush-perfused with normal saline, postfixed with 4% paraformaldehyde/0.25% glutaraldehyde in phosphate-buffered saline (pH 7.4) at 4°C for 48 h and sectioned ($40 \mu\text{m}$) with a Vibratome 2000. Sections were double-immunolabelled with a mouse monoclonal antibody against MAP-2 (Boehringer-Mannheim) and rabbit polyclonal antibodies against synaptophysin (Dako), and imaged at $0.5 \mu\text{m}$ by laser scanning confocal microscopy as described^{11,30}. For gp120 immunostaining (not shown), brains of tg mice ($n=18$) and non-tg controls ($n=18$) were processed as above or fixed in 70% ethanol/0.15 M sodium chloride and embedded in paraffin; HIV-1-infected human brain tissue which was clearly labelled with antibodies against HIV-1 gp41 was obtained post-mortem from patients with HIV encephalitis¹¹. Brain tissue homogenates were analysed on western blots using gp120 antibodies, ¹²⁵I-protein A and a Phosphorimager SF (Molecular Dynamics). Although a specific band was seen for recombinant gp120 and HL2/3-conditioned media, gp120 could not be detected in brain tissue from either tg mice or humans with HIV encephalitis (not shown).



and transgene copy numbers (range from 1 to ~80). Astroglial localization of gp120 mRNA was demonstrated by double-labelling of brain sections with a gp120-specific riboprobe and GFAP antibodies (Fig. 2a). Expression was highest in neocortex, olfactory bulb, hippocampus, tectum, selected white matter tracts and along the glia limitans (Fig. 2d).

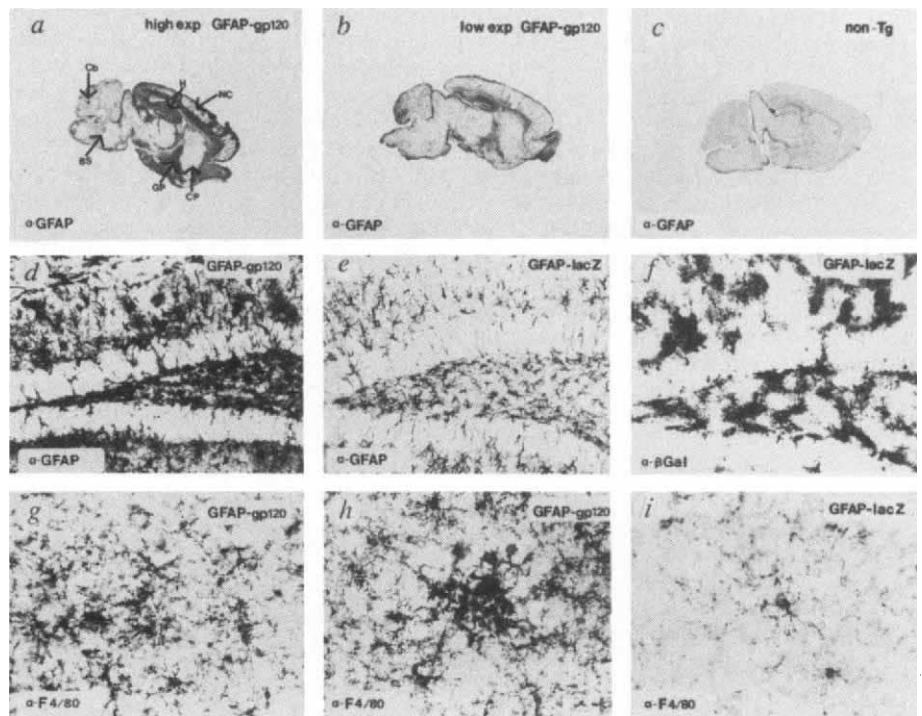
The CNS of transgenic mice was analysed for gp120 protein expression; brain tissue from AIDS patients was examined in parallel (see Fig. 2 legend). gp120 was undetectable in brain tissue from both transgenic mice and AIDS patients (not shown). This difficulty in detecting intact soluble gp120 in the CNS is well recognized and could be due to its degradation within brain tissue into neurotoxic fragments⁵. Nevertheless,

CNS expression of gp120 is strongly supported by the extensive CNS damage seen in these mice but not in controls with GFAP-driven expression of other proteins (Figs 2–4) and by the close topographical association between gp120 mRNA expression and neuropathological alterations (Figs 2 and 3).

Brains of HIV-1-infected humans display dendritic vacuolizations, loss of dendritic processes, decreases in synaptic density and loss of neuronal subpopulations^{2,11}. To determine whether CNS expression of gp120 is sufficient to induce such damage, the integrity of neurons was examined and changes quantified using laser scanning confocal microscopy of brain sections immunologically labelled for the neuronal dendritic marker MAP-2, and for a marker of presynaptic terminal boutons, syn-

FIG. 3 Glial alterations in brains of GFAP-gp120 transgenic mice. Immunostaining of sagittal brain sections (whole brain, a–c; hippocampus; d, e) with GFAP antibodies revealed normal GFAP expression by astrocytes in non-transgenic (non-tg) mice (c) and in GFAP-lacZ transgenic (tg) controls (e). In contrast, GFAP-gp120 tg mice showed a widespread increase in GFAP expression (a, b) and hypertrophy of GFAP-positive astrocytes (d). This reactive astrocytosis was most prominent in the high expressor line 2 (a). Notably, GFAP-lacZ tg controls showed normal astroglial morphology and GFAP staining (e) despite extensive GFAP-driven expression of β -galactosidase, demonstrated in f by immunostaining with antibodies against β -galactosidase. Staining of frontoparietal cortex with antibodies against the mononuclear phagocyte antigen, F4/80, revealed an abundance of highly branched microglia in GFAP-gp120 tg mice (g, h), whereas GFAP-lacZ tg controls showed a normal staining pattern (i). Occasionally, F4/80-positive cells in GFAP-gp120 tg brains were arranged in clusters reminiscent of microglial nodules (h).

METHODS. Brain sections (40 μ m) from 1.5- to 2.5-month-old tg and non-tg mice obtained as in Fig. 2 were immunolabelled as described³⁰ with the following antibodies: mouse monoclonal anti-GFAP (Boehringer-Mannheim) (a–e), rabbit polyclonal anti- β -galactosidase (Cappel, West Chester, Pennsylvania) (f), rat monoclonal anti-F4/80 (Serotec, Kidlington, Oxford, UK) (g–i). Primary antibodies



were detected with biotinylated secondary antibodies and avidin-biotin peroxidase kits (Vector) using diaminobenzidine/ H_2O_2 . Abbreviations: Cb, cerebellum; BS, brain stem; H, hippocampus; NC, neocortex; GP, globus pallidus; CP, caudate/putamen.

aptophysin (Figs 2e–h and 4a–c). Extensive vacuolizations of dendrites and a decrease in synapto-dendritic complexity were seen in high expressor lines 2 (Figs 2g and 4a–c) and 10 (not shown). Neuronal damage in the low expressor line 16 consisted primarily of dendritic vacuolizations (Figs 2h and 4c). In contrast, a normal morphology was found in non-transgenic mice (Fig. 2e) and controls with GFAP-driven expression of lacZ (Fig. 2f) or α 1-antichymotrypsin (not shown). High expressor GFAP-gp120 transgenic mice also showed a loss of large pyramidal neurons. Brain sections from five 4-month-old transgenic mice (line 2) and five non-transgenic littermates were stained with cresyl violet and cells counted as described¹². A 40% reduction in the number of neurons larger than 100 μ m² was found in the neocortex of transgenic mice ($P < 0.05$). Between groups, no significant differences were observed in counts of cells smaller than 100 μ m². The loss of specific neuronal subpopulations combined with widespread neuronal dendritic vacuolization is consistent with neuropathological findings in HIV-infected humans^{2,11}.

The effect of astroglial gp120 expression on neurons was also assessed *in vitro*. Exposure of non-transgenic hippocampal neurons to transgenic glial cultures across a physical barrier induced marked neurotoxic changes consisting of loss of dendritic arborizations (not shown) and premature death (Fig. 1d). Potential diffusible neurotoxic mediators may include gp120, biologically active gp120 fragments and/or products derived from astrocytes or microglia^{2,4,5,13,14}.

Brains of transgenic mice had a widespread reactive astrocytosis which was most prominent in lines showing high levels of gp120 expression (Figs 3a, b, d and 4d). Reactive astrocytosis includes an increase in number and size of GFAP-expressing astrocytes (compared with age-matched controls) and constitutes an important component of the brain's response to diverse

injuries¹⁵, including infection by HIV-1 (ref. 16). The increase in GFAP-positive astrocytes was associated with a marked elevation in GFAP mRNA levels (Fig. 1c). This increase in GFAP message represents a change in the level of endogenous GFAP, because the polyadenylation signal and translational stop codon at the 3' end of the gp120 sequence (Fig. 1a) prevent GFAP expression from this construct. The identical size of GFAP transcripts in transgenic and non-transgenic mice (Fig. 1c) illustrates this point. Astrocytes of transgenic mice also had a typical reactive morphology, showing increases in both cell body size and number of immunoreactive processes (Fig. 3d). Notably, mice expressing other GFAP-driven transgenes showed normal GFAP mRNA levels (Fig. 1c) and a normal distribution and morphology of GFAP-positive astrocytes (Fig. 3e).

These astroglial alterations are provocative because widespread reactive astrocytosis is one of the earliest and most consistent changes found in the HIV-1-infected human CNS^{15,16}. The reactive astrocytosis in the GFAP-gp120 model could be induced by distress signals from damaged neurons, mediators from other injury-responsive cells or more direct effects of gp120 on astroglial genes/proteins. The response of astrocytes to injury is complex, involving expression of molecules undetectable in quiescent astroglia and increased expression of other molecules found in such cells¹⁵. Although their molecular profile suggests that reactive astrocytes primarily benefit the injured CNS, certain astroglial products or disturbances of astrocyte function may contribute to the development of neuropathological alterations^{17,18}, possibly including those observed in GFAP-gp120 transgenic mice and HIV-1-infected humans.

HIV-1-infected brains frequently show microglial abnormalities¹⁹ and *in vitro* studies implicate microglia as important mediators of gp120-induced neurotoxicity^{13,14}. Immunostaining with the microglial marker F4/80 revealed that

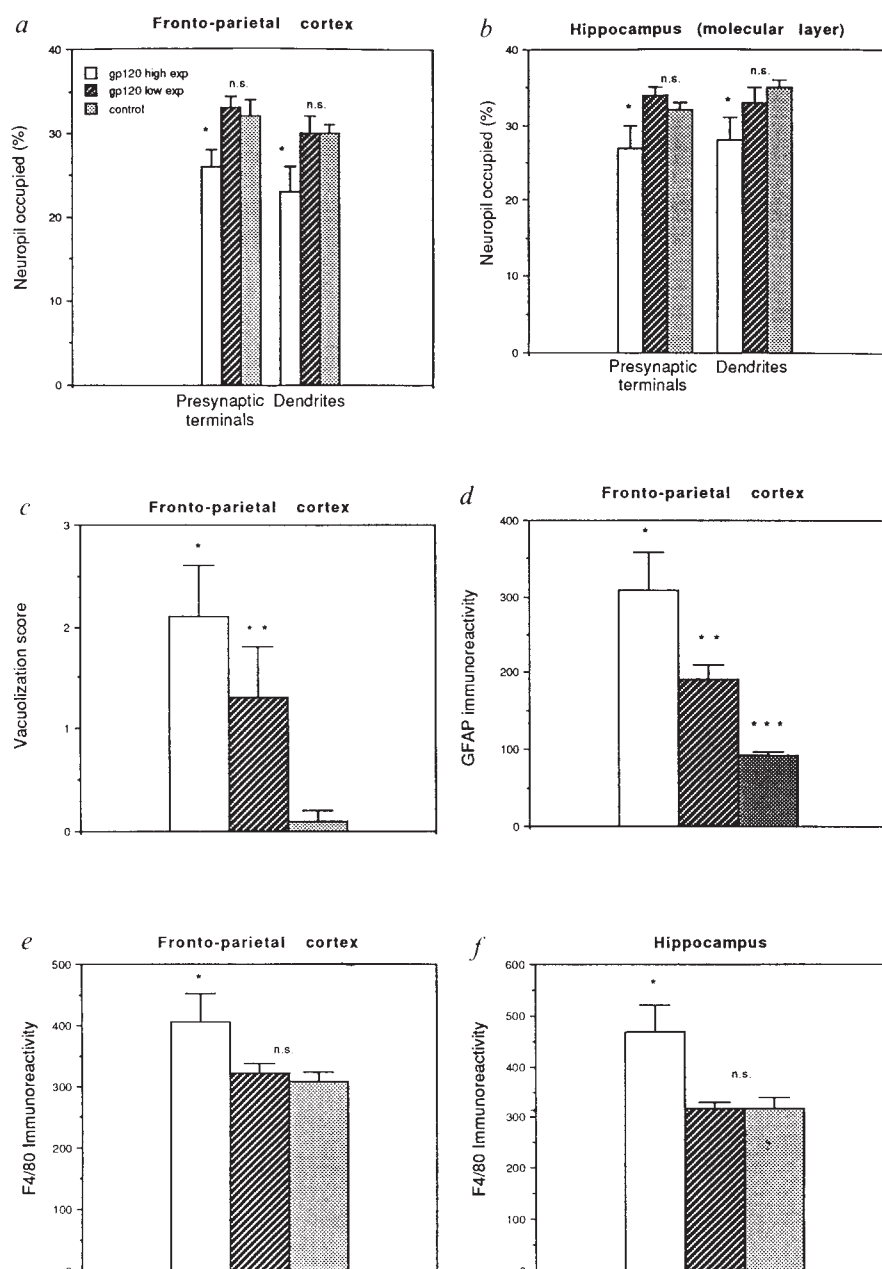
transgenic mice with high levels of gp120 expression possessed a significantly increased number of F4/80-positive cells (Figs 3g, h and 4e, f). These cells displayed increased branching and cytoplasmic distension, suggestive of microglial activation (Fig. 3g). Occasionally, microglial cells in the neocortex and hippocampus were clustered in groups (Fig. 3h) reminiscent of the microglial nodules seen in HIV encephalitis and other neurodegenerative disorders^{16,19}. Sections from mice with GFAP-driven expression of lacZ (Fig. 3i) or α 1-antichymotrypsin (not shown) exhibited a labelling pattern indistinguishable from that of non-transgenic controls (not shown).

The presence of similar neuropathological changes in separate lines of GFAP-gp120 transgenic mice indicates that they are not caused by an insertional mutation with neurodevelopmental consequences; their absence in mice expressing other GFAP-driven transgenes (Figs 1c, 2f and 3e) confirms that they do not

result nonspecifically from astroglial expression of any foreign protein. A cause-effect relationship between gp120 expression and induction of CNS damage is further supported by the positive correlation between intensity of neuropathological alterations and CNS levels of gp120 expression across different lines (Figs 1c, 2e, f, 3a, b and 4). Future experiments must address whether gp120 acts directly or indirectly in this model and identify the cellular and molecular mediators involved. Astrocytes were targeted here because astroglial expression of GFAP-driven fusion genes^{8,9,20} is an effective means of delivering soluble factors to the CNS parenchyma, and because increasing evidence implicates astrocytes in the pathogenesis of HIV-1-associated CNS damage^{2,4,18}. As monocytoic cells are the primary source of HIV-1 proteins in the human CNS², comparisons of GFAP-gp120 mice with models in which gp120 is released from macrophages/microglia will also be important.

FIG. 4 Quantification of neuropathological alterations in GFAP-gp120 transgenic mice. Transgenic (tg) mice from high expressor lines displayed statistically significant losses of dendrites and presynaptic terminals (a, b) and abundant vacuolization of apical dendrites (c). Neurons of low expressor mice showed mainly mild dendritic vacuolization (c). The intense gliosis in high expressor mice was reflected by statistically significant increases in staining for GFAP (d) and F4/80 (e, f). Lesser degrees of reactive astrocytosis were found in low expressor mice (d). a, b, * $P < 0.05$ (high versus low; high versus control); c, * $P < 0.005$ (high versus control), ** $P < 0.05$ (low versus control); d, * $P < 0.001$ (high versus low), ** $P < 0.001$ (low versus control), *** $P < 0.001$ (high versus control); e, * $P < 0.05$ (high versus low; high versus control); f, * $P < 0.01$ (high versus low; high versus control). n.s., Not significant (low versus control).

METHODS. Before processing, 1.5–2.5-month-old tg mice of line 2 (open columns) ($n=6$), line 16 (hatched columns) ($n=6$) and non-tg controls (solid columns) ($n=6$) were coded (by E.R.) to ensure objective assessment. Codes were broken only after analyses were complete. Brains were fixed, sectioned and immunolabelled for either MAP-2 (neuronal dendrites) and synaptophysin (presynaptic terminals) (a, c) or GFAP (astrocytes) (d) or F4/80 (microglia) (e, f), as in Figs 2 and 3. For each sample, three serial sections were examined using the BioRad MRC-600 laser scanning microscope^{11,30} mounted on an Axiovert Zeiss microscope. Digitized images (4 per section per case), (0.5 μ m) were computer quantified using Image 1.23 software³⁰. The area of the neuropil occupied by immunolabelled dendrites or presynaptic terminals (a, b) was expressed as a percentage of the total image area, as described previously^{11,30}. Vacuolization was graded semiquantitatively (0, none; 1, mild; 2, moderate; 3, intense) by inspection of three sections per case (c). Optical density values for relative levels of GFAP and F4/80 immunoreactivity (d–f) were obtained with the Quantimet 570C as described³⁰. Corrected optical densities were obtained by subtracting background optical densities (of sections processed without primary antibodies). Statistical analyses were conducted using the STAT VIEW II (Abacus Concepts, Berkeley, CA) software package and comparisons performed using one-factor analysis of variance. Results are expressed as mean $\pm$ s.e.m.



It should be emphasized that nervous system damage from diverse pathogenic processes can be manifested in a limited number of CNS alterations. Many of the neuropathological changes associated with HIV-1 infection can be found in other neurological diseases^{11,15,19}. Similarly, the expression of other neurotoxic factors in the CNS of transgenic mice can induce alterations that, in part, resemble those seen in the GFAP-gp120 model^{10,21}. Comparisons of GFAP-gp120 transgenic mice with models expressing other potentially pathogenic proteins should help determine the relative contributions of these factors to HIV-induced neurological disease.

In conclusion, this study demonstrates that expression of gp120 within the unmanipulated CNS is sufficient to induce pathological effects on neurons, astrocytes and microglia. The close resemblance of these changes to those found in brains of AIDS patients provides strong *in vivo* support for a causative role of gp120 in HIV-1-associated nervous system damage. This transgenic model will be valuable for assessing pharmacological approaches targeted at HIV-induced neurotoxicity and may also prove useful in the more general evaluation of neuroprotective strategies. □

6. Kimes, A. S., London, E. D., Szabo, G., Raymon, L. & Tabakoff, B. *Expl Neurol.* **112**, 224–228 (1991).
7. Glowa, J. R. et al. *Brain Res.* **570**, 49–53 (1992).
8. Mucke, L., Oldstone, M. B. A., Morris, J. C. & Nerenberg, M. I. *New Biol.* **3**, 465–474 (1991).
9. Mucke, L. et al. *Neurobiol. Aging* **13**, Suppl. 1, S101 (1992).
10. Campbell, I. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10061–10065 (1993).
11. Masliah, E. et al. *Ann. Neurol.* **32**, 321–329 (1992).
12. Wiley, C. A. et al. *Ann. Neurol.* **29**, 651–657 (1991).
13. Lipton, S. A. *NeuroReport* **3**, 913–915 (1992).
14. Giulian, D., Wendt, E., Vaca, K. & Noonan, C. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2769–2773 (1993).
15. Eddleston, M. P. & Mucke, L. *Neuroscience* **54**, 15–36 (1993).
16. Budka, H. *Brain. Path.* **1**, 163–175 (1991).
17. Abraham, C. R., Kanamaru, K. & Mucke, L. *Brain Res.* **621**, 222–232 (1993).
18. Mucke, L. & Eddleston, M. *FASEB J.* **7**, 1226–1232 (1993).
19. Dickson, D. W., Lee, S. C., Mattiace, L. A., Yen, S.-H. C. & Brosnan, C. *Glia* **7**, 75–83 (1993).
20. Mucke, L. & Rockenstein, E. M. *Transgene* **1**, 3–9 (1993).
21. Campbell, I. L., Mucke, L. & Sandberg, K. *Soc. Neurosci. Abstr.* **19**, 98.2 (1993).
22. Freed, E. O., Myers, D. J. & Risser, R. J. *Virology* **63**, 4670–4675 (1989).
23. Miura, M., Tamura, T. & Mikoshiba, K. *J. Neurochem.* **55**, 1180–1188 (1990).
24. McCune, J. M. et al. *Cell* **53**, 55–67 (1988).
25. Haass, C., Hung, A. Y. & Selkoe, D. J. *J. Neurosci.* **11**, 3783–3793 (1991).
26. Ratner, L. et al. *AIDS. Res. hum. Retroviruses* **3**, 57–69 (1987).
27. Banker, G. A. & Cowan, W. M. *Brain. Res.* **126**, 397–420 (1977).
28. Wilson, M. C. & Higgins, E. A. in *Neuromethods: Molecular Neurobiological Techniques* Vol. 16 (eds Boulton, A. A., Bake, G. B. & Campagnoni, A. T.) 239–284 (Humana, Clifton, New Jersey, 1989).
29. Masliah, E. et al. *Lab. Invest.* **66**, 285–291 (1992).
30. Masliah, E. et al. *Expl Neurol.* **113**, 131–142 (1991).

Received 4 May; accepted 5 November 1993.

1. Working Group of the American Academy of Neurology AIDS Task Force. *Neurology* **41**, 778–785 (1991).
2. Spencer, D. C. & Price, R. W. A. *Rev. Microbiol.* **46**, 655–693 (1992).
3. Schneider, J., Kaaden, O., Copeland, T. D., Oroszian, S. & Hunsmann, G. *J. gen. Virol.* **67**, 2533–2538 (1986).
4. Epstein, L. G. & Gendelman, H. E. *Ann. Neurol.* **33**, 429–436 (1993).
5. Hill, J. M., Mervis, R. F., Avidor, R., Moody, T. W. & Brenneman, D. E. *Brain. Res.* **603**, 222–223 (1993).

ACKNOWLEDGEMENTS. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: Clone pEnv (from E. O. Freed and R. Risser), polyclonal antisera to gp120 (K. Steimer, M. Phelan, M. Page and R. Thorpe), polyclonal antisera to gp160₈(HT3), gp120 from HIV₅₂ (N. Haigwood) and HIV-1 env-expressing HL2/3 cells (B. K. Felber and G. N. Pavlakis). We thank C. Evans, B. Pearce and P. Borrow for advice on *in situ* hybridization and immunocytochemistry and J. Price, E. Kothari, M. Mallory and M. Ruppe for technical assistance. This project was supported by grants from the National Institute of Mental Health (to L.M.), the National Institute of Aging (E.M. and S.T.), the National Institute of Health-AIDS (E.M.) and the Department of Health Services of the State of California (E.M.).

Superfibrinectin is a functionally distinct form of fibronectin

Alex Morla*, Zhuohua Zhang*†
& Erkki Ruoslahti*†

* Cancer Research Center, La Jolla Cancer Research Foundation, La Jolla, California 92037, USA

† Molecular Pathology Program, University of California, San Diego, School of Medicine, La Jolla, California 92037, USA

FIBRONECTIN is an extracellular matrix protein that is important in development, wound healing and tumorigenesis^{1–5}. In the blood it is dimeric, but in tissues forms disulphide crosslinked fibrils^{1,2,4}. Here we show that a fragment from the first type-III repeat of fibronectin binds to fibronectin and induces spontaneous disulphide crosslinking of the molecule into multimers of high relative molecular mass which resemble matrix fibrils. Treatment of fibronectin with this inducing fragment also converts fibronectin into a form that has greatly enhanced adhesive properties (hence the term **superfibrinectin**) and which suppresses cell migration. Whereas cells attach to fibronectin through integrins, cell attachment to superfibrinectin is mediated both by integrins and by receptors with properties distinct from those of integrins. Superfibrinectin may be closely related to the natural matrix form of fibronectin.

The first type-III repeat in fibronectin has previously been shown to play an important part in fibronectin matrix assembly^{6,7} and fibronectin–fibronectin binding⁷. We have now produced a recombinant fragment (III₁-C), modelled after the C-terminal two-thirds of the III₁ repeat, which binds to fibro-

nectin and enhances fibronectin–fibronectin binding. Treatment of purified fibronectin in solution with the III₁-C fragment resulted in the appearance of multimers of fibronectin with high relative molecular masses (M_r s) (Fig. 1a). These multimers migrated at the monomer position in samples treated with reducing agents, indicating that the multimers were held together by disulphide bonds, similar to fibronectin in the matrix *in vivo*^{8,10}. Multimer formation showed a dose dependence on III₁-C; there was no crosslinking in the presence of a negative control fragment modelled after an analogous region from the eleventh type-III repeat (fragment III₁₁; Fig. 1a). The disulphide crosslinking of fibronectin probably occurred through disulphide exchange, as free sulphhydryls were blocked in these preparations of fibronectin and the addition of iodoacetamide to the samples inhibited the crosslinking induced by III₁-C (not shown). An amino-terminal fragment of fibronectin of M_r 70,000 (70K) which is known to inhibit fibronectin matrix assembly (refs 11 and 12, and Fig. 3a) also inhibited the *in vitro* crosslinking of fibronectin induced by III₁-C (not shown), suggesting that the crosslinking may occur *in vitro* and *in vivo* by similar mechanisms.

Fibronectin becomes crosslinked *in vitro*¹³, but this crosslinking results from oxidation of free sulphhydryls, rather than through disulphide exchange, as occurs during matrix assembly¹³, and after treatment of fibronectin with III₁-C. The fibronectin multimers induced by III₁-C formed tightly packed clusters. However, the fibrillar nature of these multimers was revealed by treatment of the preformed multimers with urea, as this resulted in fibronectin fibrils reminiscent of those produced by cells (Fig. 1b, c). No multimers were detected in solutions of fibronectin treated with the III₁₁ fragment (not shown). Thus, III₁-C treatment of fibronectin results in the disulphide crosslinking of fibronectin into fibrils that are similar to the matrix fibrils produced *in vivo*.

Treatment of fibronectin with III₁-C resulted in a 10-fold increase in the ability of fibronectin to support CHO-cell adhesion and spreading (Fig. 2). The extent of cell spreading on 0.5 µg ml⁻¹ fibronectin plus III₁-C was comparable to that on

† To whom correspondence should be addressed.

10 $\mu\text{g ml}^{-1}$ fibronectin (not shown). At these coating concentrations, control experiments revealed no significant differences in the amount of fibronectin coated in the absence or presence of III₁-C (not shown). The same was observed with IMR-90 fibroblasts, T47D breast carcinoma cells (not shown) and UCLA-P3 lung carcinoma cells (Fig. 4). To activate the super-adhesiveness of fibronectin, the III₁-C fragment and fibronectin must be allowed to interact in solution before being coated onto the dish; coating with fibronectin first, followed by addition of the III₁-C fragment later, did not result in increased adhesiveness (Fig. 2). The enhanced adhesion was specific to the combination of fibronectin plus III₁-C as III₁-C did not support adhesion on its own and mixing III₁ with fibronectin had no significant effect on the ability of fibronectin to support adhesion.

The addition of III₁-C to CHO cells increased the amount of ¹²⁵I-labelled fibronectin that became incorporated into a deoxycholate-insoluble matrix (Fig. 3a). This ¹²⁵I-fibronectin was di-

sulphide crosslinked into multimers, as assessed by SDS-polyacrylamide gel electrophoresis (PAGE) analysis (not shown). Treatment of the cells with the III₁ control fragment did not affect fibronectin matrix assembly. The amino-terminal 70K fibronectin fragment inhibited matrix formation.

The ability to migrate in a 'wound' assay^{14,15} was reduced in CHO cells treated with III₁-C (Fig. 3b). Such cells migrated at ~10–20% of the rate of cells treated either with no fragment or with the control fragment III₁. This is reminiscent of cells that have been engineered to produce more fibronectin matrix by overexpressing the $\alpha 5\beta 1$ fibronectin receptor¹⁴.

In agreement with previous studies^{16–19}, dimeric fibronectin supported cell migration in a dose-dependent manner (Fig. 3c). In contrast, treatment of fibronectin with III₁-C resulted in a bell-shaped curve; at low coating concentrations the CHO cells migrated better on III₁-C-treated fibronectin than on dimeric fibronectin, whereas at higher concentrations the migration was

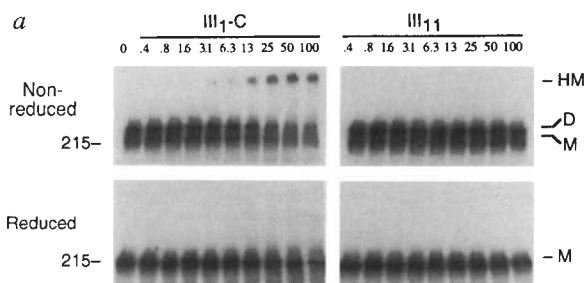
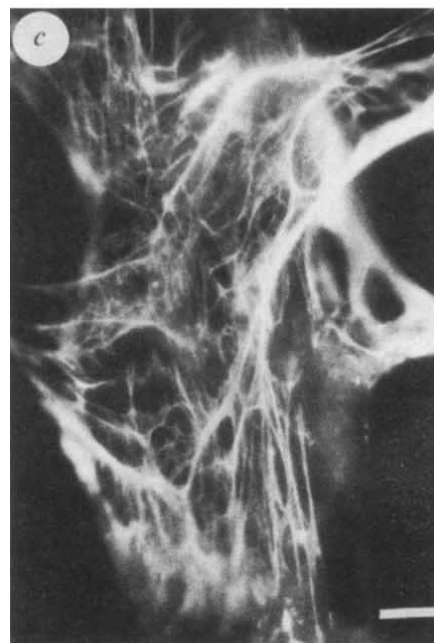


FIG. 1 *In vitro* disulphide crosslinking of fibronectin. **a**, Electrophoretic analysis of *in vitro* crosslinked fibronectin. ¹²⁵I-labelled, N-ethylmaleimide (NEM)-blocked fibronectin (1 $\mu\text{g ml}^{-1}$ in 2% BSA) was incubated with III₁-C or III₁₁ at the concentrations (μM) indicated above each lane. The samples were incubated at 37 °C for 24 h, then mixed with SDS sample buffer either with ('Reduced' samples) or without ('Non-reduced' samples) 1% 2-mercaptoethanol, followed by analysis by SDS-PAGE. The migration of a 215K marker is indicated to the left of the panels. The locations of the monomeric (M), dimeric (D) and high-M, crosslinked (HM) fibronectin are indicated to the right of the panels. **b**, Fluorescence microscopy of *in vitro* crosslinked fibronectin. Fibronectin (240 $\mu\text{g ml}^{-1}$ in 2% BSA) labelled with fluorescein isothiocyanate (FITC) was incubated with III₁-C (10 μM) at 37 °C for 20 h. Urea was added to a final concentration of 3 M and the samples were mounted for microscopy. **c**, Fluorescence microscopy of matrix fibronectin produced by IMR-90 fibroblasts in cell culture. IMR-90 cells were seeded onto glass coverslips in medium containing rhodamine-labelled fibronectin, grown for 20 h, washed with phosphate-buffered saline (PBS), then fixed with 3.7% paraformaldehyde followed by mounting of the coverslips for microscopy. In **b** and **c**, the scale bar represents 1 μm .

METHODS. The recombinant III₁-C and III₁₁ proteins were produced by polymerase chain reaction (PCR) cloning from a human placenta complementary DNA library using the following primers; 5' primer, 5'-CCG-GATCCAATGCACACAGCCATCTCA-3' and 3' primer, 5'-CCGGAT-CCAGGTGTGCTGGTGGTGGTGG-3' for III₁-C; 5' primer, 5'-CGGGATCCC-TGCCTTCAAGTTCCTCTGTT-3' and 3' primer, 5'-CGGGATCCGGTTACT-GCAGTCTGAACCAAG-3' for III₁₁. These primers amplify the regions in fibronectin encoding amino acids 600–674 (Asn-Ala-Pro-Gln...Thr-Ser-Thr-Pro) for III₁-C and 1,532–1,599 (Leu-Pro-Ser-Ser...Thr-Ala-Val-Thr) for III₁₁, according to the numbering method of Kornblihtt *et al.*²³. PCR products were cloned into the prokaryotic expression plasmid pQE-12 (Qiagen) by using the *Bam*HI sites provided by the primers. Recombinant proteins were purified on Ni-agarose columns as recommended by the manufacturer, and were >98% pure as judged by SDS-PAGE and staining of gels with Coomassie blue (not shown). The resulting proteins contain 4 amino acids at the amino termini (Met-Arg-Gly-Ser) and 10 amino acids at the carboxy termini (Gly-Ser-Arg-Ser-His-His-His-His-His) that are not normally found in fibronectin. Human plasma fibronectin was obtained from the Blood Transfusion Service of the Finnish Red Cross in Helsinki. Free sulphhydryl groups in fibronectin were blocked with NEM as described previously¹³. The NEM-blocked fibronectin was then labelled with ¹²⁵I as described⁷.



suppressed on III₁-C-treated fibronectin (Fig. 3c). Cells adhered to and spread very well on high concentrations of III₁-C-treated fibronectin (Fig. 2), suggesting that the adhesion to this substrate may be too tight for migration. No migration was found on filters that were coated with the III₁-C fragment alone (Fig. 3c). The reduced migration of cells on III₁-C-treated fibronectin could not be explained by differences in coating efficiency in the presence of the III₁-C fragment, as the coating efficiencies were similar for the two forms of fibronectin (not shown).

FIG. 2 Cell adhesion to fibronectin that has been treated with the III₁-C fragment. Fibronectin (FN) at various concentrations was coated onto plastic dishes either in the absence (●) or presence of 1 μ M III₁₁ per μ g ml⁻¹ fibronectin (○) or 1 μ M III₁-C per μ g ml⁻¹ fibronectin (▲). Alternatively, wells were coated with III₁-C alone (■) or first coated with various concentrations of fibronectin followed by blocking with 0.5% BSA, then addition of III₁-C (□). After coating, all wells were blocked with 0.5% BSA, then C11 cells (variant CHO cells described in ref. 13) were allowed to adhere for 30 min at 37 °C. They were then fixed, stained and the amount of cell adhesion was measured.

METHODS. Cell attachment to coated fibronectin in the presence or absence of added III₁-C or III₁₁ was performed as described previously^{14,24}. A concentration of 1 μ M III₁-C per μ g ml⁻¹ fibronectin resulted in optimal adhesion (not shown), and was therefore used for both III₁-C and III₁₁. All protein solutions were in PBS. Cell adhesion was quantified by staining cells with 0.5% crystal violet in 20% methanol, washing with water, then eluting the dye with 0.1 M sodium citrate, pH 4.2, 50% ethanol and measuring the absorbance at 600 nm.

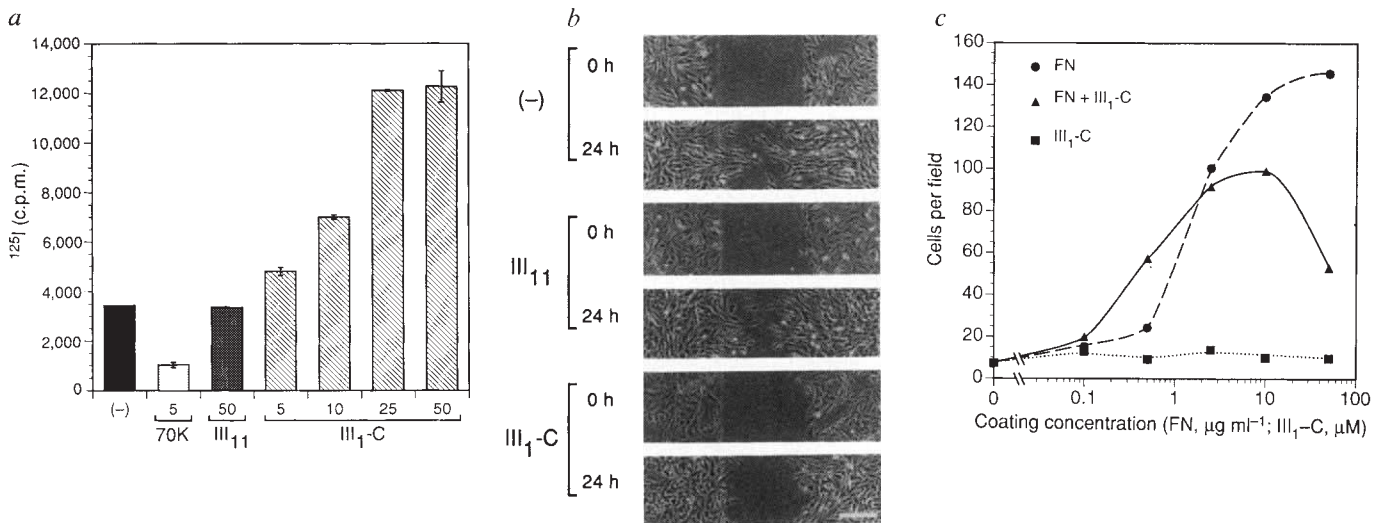
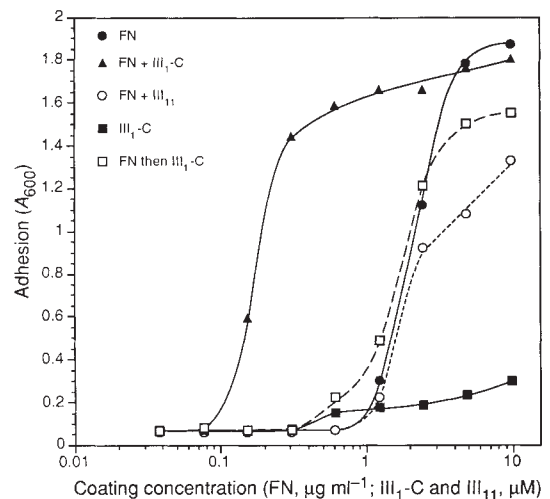


FIG. 3 Effect of the III₁-C fragment on ¹²⁵I-fibronectin deposition and cell migration. **a**, C11 cells were grown to subconfluence and incubated for 24 h with ¹²⁵I-fibronectin in the presence of various concentrations of unlabelled fragment III₁-C, fragment III₁₁ or amino-terminal 70K fragment, as indicated. The concentrations (μM) of the fragments used in the assays are indicated below each column. Cells were washed followed by quantification of the amount of ¹²⁵I-fibronectin found in a deoxycholate-insoluble pool. Each data point is the average of duplicate determinations; error bars indicate the variation among duplicates. **b**, Subconfluent monolayers of C11 cells were wounded at time 0 and photographed. Cells were allowed to migrate into the vacated area for 24 h, then photographed again. The figure shows representative fields of cells that had been grown in no fibronectin fragment or 50 μ M III₁₁ or 50 μ M III₁-C as indicated. Scale bar, 200 μ m. **c**, Boyden chamber filters were coated with various concentrations of fibronectin in the absence (●) or presence (▲) of the III₁-C fragment at 1 μ M III₁-C per μ g ml⁻¹ of fibronectin. Alternatively, filters were coated with III₁-C fragment only (■). C11 cells were plated onto the upper chamber and allowed to migrate into the lower chamber for 4 h. They were fixed and stained, then the cells on the upper surface were scraped off the filters

and those on the lower surface were counted. All data points are the average of three randomly chosen microscopic fields; variation between samples was less than 10%. **METHODS.** Matrix assembly assays were performed by using ¹²⁵I-fibronectin as described previously^{7,11}. Cells were grown in 24-well dishes in complete culture medium for 48 h before the addition of ¹²⁵I-fibronectin plus either III₁-C or III₁₁ as described above. Cell migration assays were performed as described previously^{14,15}. Cells were grown in 35-mm dishes in culture media plus either no additions or fragment III₁-C or fragment III₁₁ for 48 h. Cell cultures were then wounded with a pipette tip. Photographs of marked areas on the dishes were taken immediately after wounding and then again 24 h later. Cell migration in the wound assay was quantified by comparing matched sets of 0-h and 24-h photographs and measuring the distance migrated by the majority of cells. Migration in a Boyden chamber was measured as described previously²⁴. Lower chambers were filled with 0.1% BSA, 25 ng ml⁻¹ platelet-derived growth factor (PDGF) in α -minimum Eagle's medium. Cells (5×10^4) were added to the upper chambers in the same medium except without PDGF. They were allowed to migrate for 4 h at 37 °C.

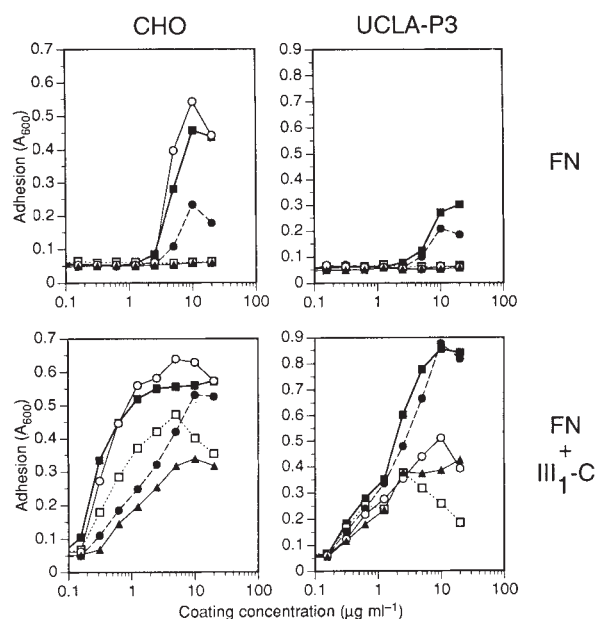


FIG. 4 Binding of cells to III₁-C-treated fibronectin through more than one class of receptors. CHO and UCLA-P3 cells were allowed to adhere to various concentrations of fibronectin or III₁-C-treated fibronectin for 1 h at 37 °C in the absence (■) or presence of 10 mM EDTA (□), 100 μg ml⁻¹ anti-fibronectin receptor antibodies (●), 100 μg ml⁻¹ anti-vitronectin receptor antibodies (○) or a combination of both antibodies (each at 100 μg ml⁻¹) plus 10 mM EDTA (▲). Note that CHO cells use the α5β1 integrin as their main fibronectin receptor¹⁹ whereas UCLA-P3 cells use αvβ5 as a fibronectin receptor^{20,21}. METHODS. Cell adhesion assays were performed as described in Fig. 2 legend. Antibodies used were rabbit polyclonal anti-fibronectin receptor (to block α5β1; ref. 25) and anti-vitronectin receptor (to block αvβ5; ref. 26).

to III₁-C-treated fibronectin was only partially inhibited by EDTA (Fig. 4). For UCLA-P3 cells, the integrin-mediated binding component on III₁-C-treated fibronectin was higher than on regular fibronectin, indicating that integrin binding was enhanced by III₁-C treatment. Thus, the superadhesiveness of cells on III₁-C-treated fibronectin may be due to binding of the cells through more than one class of receptors: integrins and a family of receptors not related to the integrins. This is in sharp contrast to the binding of cells to regular fibronectin, which is mediated entirely by integrins. The identity of the second class of receptors is unknown.

As attempts to demonstrate the presence of a fibronectin fragment equivalent to III₁-C in cell cultures or in mouse tissues have been unsuccessful, we suspect that *in vivo* this activity resides within the intact fibronectin molecule and that it may

become unmasked at the appropriate time during matrix assembly, perhaps as a result of binding of fibronectin to the α5β1 integrin.

Superfibrinectin may be closely related to the *in vivo* matrix form of fibronectin: both are disulphide crosslinked into fibrils through a disulphide exchange mechanism, the crosslinking of both is inhibited by an amino-terminal 70K fragment of fibronectin, and cells migrate less when they are either in the presence of superfibrinectin or engineered to produce more fibronectin matrix. Because of the similarities between superfibrinectin and the fibronectin matrix, the interaction of cells with superfibrinectin may more accurately approximate the interaction of cells with a fibronectin matrix *in vivo*. Other extracellular matrix proteins may undergo similar transitions on assembly into a matrix. □

Received 29 April; accepted 5 November 1993.

- McDonald, J. A. A. *Rev. Cell Biol.* **4**, 183–207 (1988).
- Ruoslahti, E. A. *Rev. Biochem.* **57**, 375–413 (1988).
- Ruoslahti, E. & Giancotti, F. G. *Cancer Cells* **1**, 119–126 (1989).
- Hynes, R. O. *Fibronectins* (Springer, New York, 1990).
- Mosher, D. F., Sotile, J., Wu, C. & McDonald, J. A. *Curr. Opin. Cell Biol.* **4**, 810–818 (1992).
- Chernousov, M. A., Fogarty, F. J., Kotliansky, V. E. & Mosher, D. F. *J. Biol. Chem.* **266**, 10851–10858 (1991).
- Morla, A. & Ruoslahti, E. *J. Cell Biol.* **118**, 421–429 (1992).
- Hynes, R. O. & Destree, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2855–2859 (1977).
- Keski-Oja, J., Mosher, D. F. & Vaheri, A. *Biochem. biophys. Res. Commun.* **74**, 699–706 (1977).
- Ali, I. U. & Hynes, R. O. *Biochim. biophys. Acta* **510**, 140–150 (1978).
- McKeown-Longo, P. J. & Mosher, D. F. *J. Cell Biol.* **100**, 364–374 (1985).
- McDonald, J. A. et al. *J. Biol. Chem.* **262**, 2957–2967 (1987).
- Mosher, D. F. & Johnson, R. B. *J. Biol. Chem.* **258**, 6595–6601 (1983).
- Giancotti, F. G. & Ruoslahti, E. *Cell* **60**, 849–859 (1990).
- Goodman, S. L., Vollmers, H. P. & Birchmeier, W. *Cell* **41**, 1029–1038 (1985).
- Furcht, L. T., McCarthy, J. B., Salm, S. L., Basara, M. L. & Enenstein, J. *Ciba Fdn Symp.* **108**, 130–145 (1984).

- Lacovara, J., Cramer, E. B. & Quigley, J. P. *Cancer Res.* **44**, 1657–1663 (1984).
- McCarthy, J. B., Basara, M. L., Salm, S. L., Sas, D. F. & Furcht, L. T. *Cancer Metastasis Rev.* **4**, 125–152 (1985).
- Bauer, J. S., Schreiner, C. L., Giancotti, F. G., Ruoslahti, E. & Juliano, R. L. *J. Cell Biol.* **116**, 477–487 (1992).
- Cheresh, D. A., Smith, J. W., Cooper, H. M. & Quaranta, V. *Cell* **57**, 59–69 (1989).
- Pasqualini, R., Bodorova, J., Ye, S. & Hemier, M. E. *J. Cell Sci.* **105**, 101–111 (1993).
- Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491–497 (1987).
- Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. *EMBO J.* **4**, 1755–1759 (1985).
- Zhang, Z. et al. *J. Cell Biol.* **122**, 235–242 (1993).
- Argaves, W. S. et al. *J. Cell Biol.* **105**, 1183–1190 (1987).
- Suzuki, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8614–8618 (1986).

ACKNOWLEDGEMENTS. We thank E. Engvall and B. Wang for critical review of the manuscript. This work was supported by grants from the National Cancer Institute to E.R.; A.M. is supported by a Damon Runyon-Walter Winchell Cancer Research Fund Fellowship.

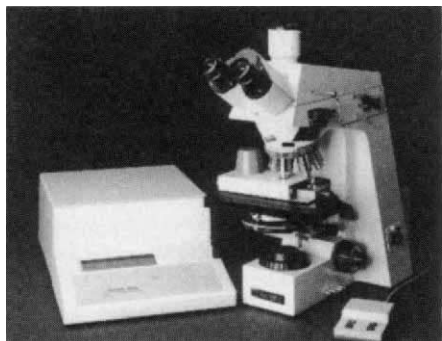
Ringling the changes

New products for the New Year include a video tracking, motion analysis and behaviour recognition system, a new system for measuring the molecular binding of biological molecules and an HPLC system with mass spectral compound identification.

EthoVISION is a new system for the **automatic recording of activity, movement and interactions of animals** available from Noldus Information Technology (*Reader Service No. 100*). This complete video tracking, motion analysis and behaviour recognition system, which can be used to analyse the interactions between multiple animals observed in a single enclosure, consists of a personal computer with 486 processor, colour VGA monitor, keyboard, mouse and DOS operating system; high-resolution video camera and video monitor; frame grabber; EthoTrack software; and various cables and connectors. EthoVision can be used to automate many standard tests carried out with rodents, insects and other animals.

The new **Beacon fluorescence polarization system** from PanVera is designed to measure the molecular binding of a wide range of biological molecules including proteins, nucleic acids, carbohydrates, lipids and drugs (*Reader Service No. 101*). It is said to be particularly useful for performing fluorescence immunoassays, DNA/protein binding, protease assays, DNA hybridization and quantification, lipid membrane association and fluidity, nuclease assays, receptor/ligand binding, serum drug measurements and protein/protein binding. Because the instrument measures binding values directly in solution, the need for any filtration, gel shift assays, immobilization or radioisotopes is eliminated. Although the Beacon system can be operated independently with a paper read-out, it comes with a PC-based software package for data collection and analysis.

Carl Zeiss announces that it now has exclusive rights to sell the Mettler FP900 thermosystem for thermal microscopy applications. The system offers a choice of two **microscope hot stages**, FP82 HT and FP84 HT, for investigating and documenting the temperature dependence of optically observ-



Mettler's hot stages — now available from Carl Zeiss.

able processes such as phase transition (*Reader Service No. 102*). The FP84 HT will also couple thermal microscopy with the DSC thermal analysis methods so that thermal transition data can be collected at the same time. Both hot stages are controlled by the FP90 central processor, which provides programmable temperature control, data acquisition and printer output functions. The stages integrate with the full range of Zeiss microscopes.

■ Molecular means

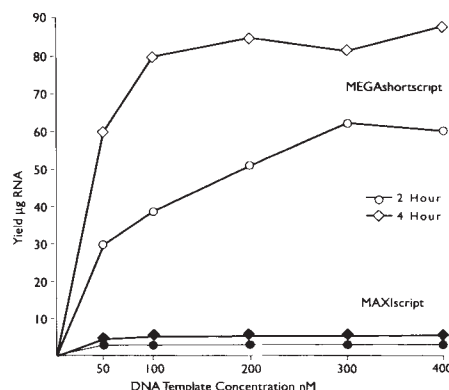
With the introduction of its **PR 1000 gel/membrane processor**, Hoefer says that it has automated the 10 to 15 steps required for chemiluminescent and colorimetric detection of nucleic acids and proteins on Southern, western, sequencing and slot blots



Hoefer's PR 1000 gel/membrane processor automates nonradioactive detection on blots.

(*Reader Service No. 103*). The device controls the selection of reagents, volume delivery, incubation time and reagent removal for each step of the protocol. The company says that most protocols can be adapted for use with the PR 1000: each procedure can have as many as eight reagents, 20 steps and step incubation times up to 999 minutes. Nine separate protocols can be stored in nonvolatile memory for quick recall. The PR 1000 accommodates membranes from small strips to full-sized sequencing gels (33 × 55 cm) with a choice of five development chambers. In the development chambers, membranes are exposed to solutions with a gentle rolling and rocking motion that is designed to increase the sensitivity of detection and reduce background.

Ambion's new T7 MEGAscript kit is designed to give **high yields of small RNAs from short templates** (< 300 bases)



Increasing amounts of a 125-base PCR product template containing a T7 promoter were transcribed in either a standard (20 µl) MEGAscript reaction or a large-scale (50 µl) MAXIsScript reaction. The yields with the former were found to be 10–20-times greater.

(*Reader Service No. 104*). The procedure uses a high polymerase enzyme mix to maximize yields of small transcripts. In addition, the use of the company's MEGAscript *in vitro* transcription technology is said to result in yields of RNA 10–20-fold greater than that in conventional *in vitro* transcription reactions. The kits, which are designed to work with oligonucleotide or PCR product templates as well as standard plasmid templates, include T7 enzyme mix which contains RNase inhibitor, ATP, CTP, GTP and UTP solutions, 10 × transcription buffer, RNase-free DNase I, linearized DNA control template, ammonium acetate stop solution, RNase-free distilled H₂O and gel-loading buffer.

Fluka's new **RNA isolation kit**, which uses a single-step extraction procedure (guanidine thiocyanate/phenol/chloroform extraction), can be used to isolate undegraded total RNA in less than 4 h (*Reader Service No. 105*). The procedure is said to enable the isolation of total RNA from at least 7 g of tissue. The individual components of the kit all bear Fluka's new quality-grade label MicroSelect for molecular biology, which, according to Fluka, means that they have been subjected to rigorous quality-control testing and are, in addition, free of DNases, RNases, proteases and phosphatases.

Whatman LabSales introduces a new dot blot technology in the AccuMax and DNA Kwik Screen cartridge — a lightweight, portable and **quantitative system for nucleotide detection** (*Reader Service No. 106*).

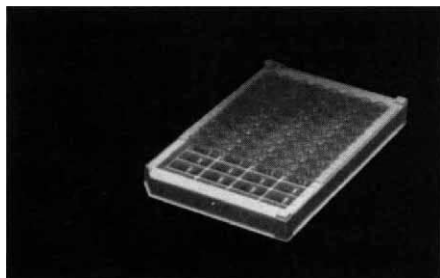
NEW ON THE MARKET

Using the membrane as the reaction site, the AccuMax produces an LED display of the reaction. Lit'l Jaws cuts your choice of membrane to size and the Red Snapper is said to make placement into DNA Kwik Screen cartridge a snap. For variable speed vacuum pressure, the DNA Kwik Screen cartridge mounts directly onto the VacMax.

The Dynabeads mRNA Direct kit from Dynal is designed for the rapid **isolation of purified intact mRNA directly from crude samples** (*Reader Service No. 107*). The methodology uses superparamagnetic Dynabeads Oligo (dT)₂₅, eliminating the need for spin columns, filtration, organic reagents and ethanol precipitation. Dynal says that a minute amount of starting material is required. Purification is performed in a single tube and enzymatic downstream applications are not inhibited by the presence of the Dynabeads. The protocol for the kit involves cell lysis followed by solid-phase hybridization. An external magnet is applied to wash the mRNA-Dynabead complex. The purified mRNA can be used directly in various applications including RT-PCR, solid-phase cDNA library, S1 nuclease analysis, ribonuclease protection, primer extension, dot and slot hybridization, *in vitro* translation experiments, subtractive hybridization and northern blotting.

■ Radioactive round-up

Wallac's new ScintiStrip 96-well **microtitration plate incorporates scintillant as part of the plastic** (*Reader Service No. 108*). The plates can be easily separated into eight-well strips and single wells and are suitable for all solid-phase binding assays that use radioactive tracers — RIAs/IRMA's,



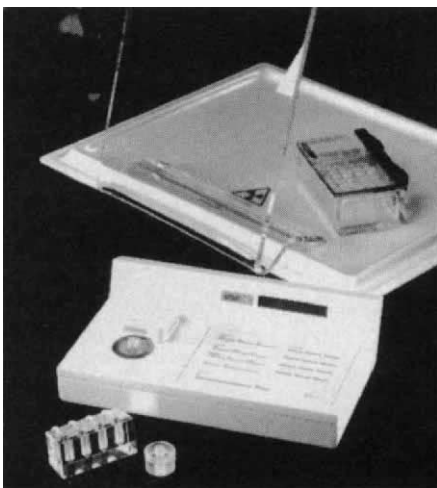
Hardy ScintiStrip plates and strips.

DNA/RNA hybridizations, membrane receptor studies, enzyme/receptor kinetics, kinetics involving titrated compounds, and any binding experiment using a labelled biotinylated compound. The ScintiStrip plates and strips can be used with microplate counters such as Wallac's new MicroBeta Plus as well as conventional beta counters. They are available in a high-binding uncoated version, as well as versions coated with anti-rabbit IgG, anti-mouse IgG and streptavidin, or with custom coatings.

Exscint II from National Diagnostics is an **extractive scintillation solution** formulated to extract the radiolabelled sample

from aqueous solution and solubilize it in the organic phase of the liquid scintillation solution (*Reader Service No. 109*). The addition of 1 ml of aqueous sample in 10 ml of Exscint II results in the formation of two transparent phases: the lower phase contains the water and other impurities, the upper one the radiolabelled sample, which can then be counted (without quenching and the effects that this can have on counting efficiency). This procedure is said to be suitable for any sample, such as steroids, having a greater solubility in aromatic solvents than in water.

For non-destructive, **quantitative monitoring of radioactive samples** in the laboratory in, for example, probe labelling reactions, Scotlab has developed the Easicount range of benchtop radiation counters (*Reader*



Easicount — for non-destructive, quantitative, benchtop radiation counting.

Service No. 110). The device can be used to monitor samples labelled with ³²P, ¹²⁵I, ¹⁴C and ³⁵S without the need for scintillation cocktails and without the generation of liquid radioactive waste. Samples can be counted directly in standard laboratory vials, such as test tubes and microtubes, using a wide range of removable sample holders.

■ Peptide/protein paraphernalia

MatchMaker **protein structure prediction software** from Tripos allows researchers to predict three-dimensional protein models by including interactions with neighbouring residues (*Reader Service No. 111*). The program, which is based on an algorithm for inverse protein folding that was developed by Jeff Skolnick and colleagues at Scripps Institute, is said to provide scientists with the ability to recognize entire protein folds and families from sequences of globular proteins using a large database of known structures. Tripos says that with MatchMaker, it is not necessary to depend solely on the sequence homology of known protein structures. In addition, the program bypasses the consideration of secondary structure to include specific local interactions

among neighbouring residues. MatchMaker is available on Silicon Graphics' workstations, and it is available stand-alone or as a Tripos SYBYL molecular analysis option.

With the new **DEPRO peptide deprotection kits**, Sigma is offering an alternative to liquid hydrogen fluoride in solid-phase peptide synthesis (*Reader Service No. 112*). The kits contain the reagents necessary for the cleavage of a peptide from 1–2 grams of resin (Merrifield resin ester, MBHA resin amide or PAM resin ester) and the removal of side-chain protecting groups normally removed by hydrogen fluoride. The kit utilizes a mild new reagent, trimethylsilyl trifluoromethanesulfonate, which eliminates potential exposure to hydrogen fluoride.

■ Chromatography/MS

An assortment of Vantage-S **columns for preparative biochromatography** are now available from Amicon (*Reader Service No. 113*). Design features include an easy-to-use



Vantage-S preparative biochromatography columns from Amicon.

adjuster mechanism and the sanitary distribution cells and seals that are said to ensure column hygiene. A wide range of column diameters is offered and bed volumes range from 1.5 to 25 litres.

Integrity is the new **HPLC system with integrated mass detection** from Waters chromatography division of Millipore designed with chromatographers in mind (*Reader Service No. 114*). System components include a Waters 616 solvent delivery system which provides four-solvent gradient capability for narrow-bore chromatographic techniques; a Waters 996 photodiode array detector designed to complement mass spectral detection by verifying that individual compounds have been separated and have not coeluted, and for corroborating the identification of compounds; a Waters ThermoBeam mass detector; and the Millennium 2010 chromatogra-

NATURE · VOL 367 · 13 JANUARY 1994



Waters' Integrity HPLC system with Integral mass spectral compound identification.

phy manager data collection, processing and reporting system. The system's mass detector provides library searchable mass spectra for the identification of sample constituents. The company says that it can identify a wide range of organic molecules of up to 1,000 atomic mass units.

The Prism stable isotope mass spectrometer from Fisons Instruments has a new low profile ergonomic design incorporating several new features — multi-tasking electronics and software, turbomolecular pumping and an adjustable multiple collector for stable isotope mass spectrometry



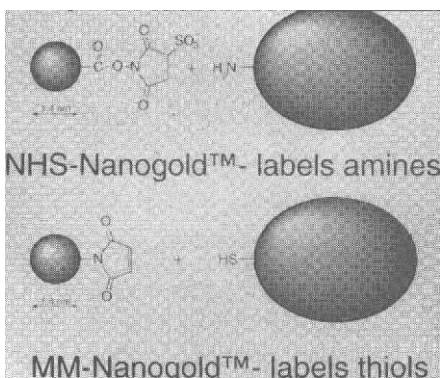
Prism — a stable isotope mass spec.

(Reader Service No. 115). The rotated focal plane of the Prism's ion optics enables the incorporation of a new adjustable multiple collector with four high-performance Faraday collectors. Fisons says that whereas no adjustment is required for routine analysis of CO_2 or N_2 , the micrometer-controlled collector is easily adjusted for future research needs, providing the flexibility needed to analyse from Ar to SF_6 . Vacuum quality is said to be improved by the inclusion of new ultra-clean turbomolecular pumping throughout. Turbo pumps are fitted at both the source and collector, with the option of differential pumping via an additional turbo pump for those researchers who want to optimize the performance of the vacuum and enhance abundance sensitivity. For hydrogen analysis, the new Prism retains its 'two-mass-spectrometers-in-one' concept incorporating a second computer-controlled electromagnet to deflect masses 2 and 3 to a new hydrogen dual collector. Multi-tasking software enables the user to run user-defined programs to control accessories and peripherals while, at the same time, performing

data acquisition and reprocessing. Prism interfaces with Fisons' sample preparation systems.

■ Assorted reagents

Nanoprobes announces two new lines of **molecular gold labelling reagents** both of which make use of the company's small (1.4 nm diameter) Nanogold gold particles (Reader Service No. 116). The first includes monomaleimido-Nanogold, which reacts with thiols, and mono-*N*-hydroxysuccinimide ester-Nanogold, which couples to amines. Molecules that can be labelled include antibodies, other proteins, peptides, lipids, carbohydrates, RNA and DNA. These gold conjugates can then be used as probes for electron or light microscopy, cytohistochemistry, diagnostic applications, blots,

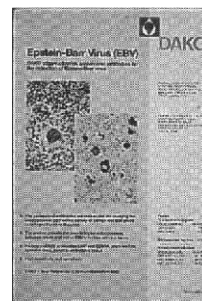


Nanoprobes' new gold labelling reagents.

westerns or gels. Silver enhancement enlarges the gold particles to 10–100 nm for easy detection by microscopy and amplifies the signal so that 10^{-19} moles can be detected on blots by eye. Nanoprobes has also developed small gold particles covalently linked to various lipids. These reagents can be used to visualize the lipid phase directly by electron and light microscopy. At present, the company is supplying dipalmitoylphosphatidylethanolamine-Nanogold. Other lipids are available on request.

Epstein-Barr virus (EBV), a human herpesvirus carried in a latent form by approximately 80 per cent of the adult population, is associated with carcinomas

and lymphomas including Hodgkin's disease. Dako is offering **oligonucleotides and antibodies to EBV** for *in situ* hybridization and immunohistochemistry (Reader Service No. 117). The panel of probes and antibodies, which can be used on formalin-fixed, paraffin-embedded tissue sections, makes it possible to distinguish between latent and active EBV infections.



EBV oligos and antibodies.

Ancell is offering **anti-CD80 (B7/BB1)** in three forms — purified, biotinylated and conjugated with fluorescein isothiocyanate (Reader Service No. 118). This monoclonal antibody should provide a useful research tool for T-cell/APC interaction. The company's 1994 catalogue also includes 20 other monoclonals to various human leukocyte cell surface molecules. New T-cell/B-cell monoclonals and Ancell's newest research tool — **soluble human leukocyte cell surface molecules** — will be available during 1994.

■ Catalogues/software

Over 2,000 **immunodiagnostic kits and reagents** — over 200 of which are new — are featured in the new product catalogue from The Binding Site (Reader Service No. 119). New kits include IgG subclass assays for use on nephelometric and turbidimetric analysers, ELISAs for thyroid microsomal and thyroglobulin autoantibodies, RID assays for human CSF and urinary proteins, CRP latex-enhanced assay for use on the BNA, slide latex tests for the identification of *Staphylococcus aureus*, slide latex tests for Streptococcal grouping, and stained bacterial suspensions. The company has also expanded its range of polyclonal and monoclonal antisera to both human and animal proteins. New F(ab)_2 antibody fractions as well as whole antibodies and conjugates have been added to the affinity-purified range and a Fab monomer conjugated with fluorescein isothiocyanate for use in flow cytometry is available for the first time.



Immunodiagnostic paraphernalia.

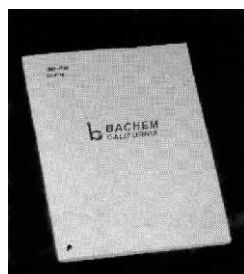
Calbiochem-Novabiochem has just published its 1994–95 **catalogue of biochemicals and immunochemicals** (Reader Service No. 120). The catalogue includes about 3,000 products (antibodies,

inhibitors, ionophores, proteins and products for signal transduction research), of which 300 are new. Along with a detailed product description, the catalogue provides application notes, CAS numbers, Merck-Index, molecular weight, hazard symbols, literature references and information concerning solubility.

Over 315 pages of laboratory instrumentation, equipment and supplies for the research laboratory are featured in the new 1994 annual **laboratory instrumentation catalogue** from Whatman LabSales (*Reader Service No. 121*). The new catalogue includes sections for research equipment, laboratory instrumentation, safety products, gen-

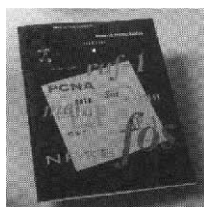
eral laboratory supplies and an expanded section for biotechnical applications containing a complete listing of Whatman filtration and chromatography products. New products range from blotters and gel dryers to glassware and plasticware sections from Corning and Nalge. In addition, the catalogue offers a line of pH meters and electrodes, conductivity equipment, mixers and stirrers.

Copies of Bachem's 1993-1994 **catalogue of bioactive peptides**, synthetic intermediates and recombinant growth factors and cytokines are now available (*Reader Service No. 122*). The catalogue features hundreds of new products, including endothelin antagonists, B-amyloid peptides, specific neurotransmitter antagonists and numerous substrates for studying protein kinase and phosphatase regulation.



Bioactive peptides and synthetic intermediates.

A new catalogue is available from Santa Cruz Biotechnology offering a broad range of **products for signal transduction research** (*Reader Service No. 123*). Product development highlights include cell cycle proteins (a line of cyclin A, B, D1, D2, D3 and E antibodies, as well as antibodies specific for each of eight cyclin-dependent

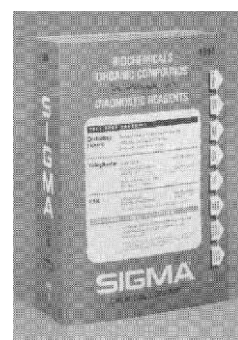


Tools for signal transduction research.

kinases); basal transcription factors (TFIIB, TFIID-TBP, TFIIE- α and β , TFIIF (RAP30 and RAP74) and USF antibodies, purified TFIIB and TFIID-TBP proteins); regulatory transcription factors (antibodies to AP-1, Sp1, AP-2, CREB, ATF 1-4, egr 1-3, NF κ B p65 and p50, rel, I κ B, SRE-ZBP, C/EBP (α , β and γ), myc, max, mad, E2F-1, oct 1 and 2, and a line of gel-shift oligonucleotides); signal transduction intermediates (antibodies to ras (H, N and K), raf, MAP kinases, MEK, MEK kinases, GRB2, Sos, rap, rho, rac, GAP, ras-GRF, db1 and vav, and protein kinase substrate and inhibitor peptides); and cell membrane receptors (antibodies to EGF(R), neu, trk (A, B and C), met, FGF(R), c-fms/CSF-1, PDGF(R), TGF β TYPE II (R)).

Sigma's Chemical's new 1994 **catalogue** includes over 33,000 product listings for biochemical and organic compounds, diagnostic reagents, and laboratory equipment and supplies (*Reader Service No. 124*). Several new product areas are represented —

chromatography reagents, filtration supplies and SigmaUltra high-purity chemicals. Sigma's specialty product groups include enzymes, peptides, immunochemicals, and reagents developed for molecular biology and tissue culture.



Sigma's supplies.

SciWords for Agriculture, the latest **spell-checker** in the SciWords series available from Pool, Heller and Milne, contains over 72,000 agricultural-related words including names of plant and animal species, and micro-organisms, chemical and pesticide trade names, and common acronyms used in technical writing (*Reader Service No. 125*). All SciWords spell checkers work with standard WordPerfect (DOS, Windows, Macintosh) or Microsoft Word (Windows, Macintosh) dictionaries. The company says they integrate easily and are activated automatically every time a spell check is requested. SciWords for chemistry contains over 75,000 chemical names and technical terms used in chemistry, physics and biology, with special emphasis on pharmaceuticals. The spell-checker programs cost US\$60; volume discounts, academic discounts, LAN and site licences are available. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal.

Erratum

Nature 366, 187 (1993)

In the above product review (*Reader Service No. 114*), the image analysis software program for electrophoresis applications was incorrectly called the Phoretix 2D program. It should have read Phoretix 1D.

Correction

H. Leirs & L. De Bruyn

Nature 366, 183-184 (1993)

In a recent product review, the unintentional omission of three words suggested that the bibliographic software EndNote Plus (Niles & Associates, Inc., Berkeley, California) does not support keywords. This is wrong, as EndNote Plus has a special keywords field that can store up to 32,000 characters. The software can import keywords from a variety of online databases such as Medline, BIOSIS and Chem Abstracts. When searching for references, users of EndNote Plus can search the keywords field as well as any other field. The keywords field can be indexed for faster searching. Although Neal regrets that the program does not maintain a list of keywords, he found the program highly satisfactory.

1. Neal, P. R. *Bioscience* 43, 44-51 (1993).

ADVERTISEMENTS

DNA SEQUENCING

Using an ABI 373 Automatic Sequencing System

£30/sample

- PCR products - single stranded DNA
- M13 inserts - double stranded DNA
- plasmid inserts - solid phase sequencing
- 300-400 bases of sequence

Please contact us for more information and a preparation protocol

Molecular Medicine Unit

King's College School of Medicine & Dentistry
Tel: 071-326 3126 Fax: 071-733 3877

Reader Service No.10

OLIGONUCLEOTIDES

Synthesised, purified and shipped in 3 days

£2.75 per base

(All inclusive)

0.2 μ M scale (approx. 500 μ g)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry
Tel: 071-326 3126 Fax: 071-733 3877

Reader Service No.11

Antibodies and Antigens for Extracellular Matrix Research

- ✓ Collagens
- ✓ Fibronectin
- ✓ Fibronectin Receptors
- ✓ Integrins
- ✓ Laminins
- ✓ Merosin
- ✓ Proteoglycans
- ✓ Tenascin
- ✓ Vitronectin
- ✓ Vitronectin Receptors

For a complete listing of our products please call, fax or write us today.

800-437-7500



CHEMICON International, Inc.

28835 Single Oak Drive, Temecula, CA 92590 USA
Phone 909-676-8080 • Fax 909-676-9209

Reader Service No.13

Genetics of longevity and more trinucleotide repeats

The first issue of *Nature Genetics* this year includes the first search for alleles associated with longevity outside the human HLA complex and the seventh neurodegenerative disease associated with a repetitive DNA triplet sequence.

How to begin on the enumeration of the genetic components of longevity? That is the task taken up by a group from the Centre d'Etude du Polymorphisme Humain (CEPH) at Paris in the current issue of *Nature Genetics* (F. Schächter *et al.* *Nature Genet.* 6, 29; 1994). The starting point must be a group of people who have lived a long time, and whose genetic constitution may be compared with that of younger people. The CEPH group boasts of having found no fewer than 338 centenarians in metropolitan France (of whom 87 per cent are, not surprisingly, women). The control group is drawn from CEPH's database of pedigrees and comprises people aged between 20 and 70 years.

The underlying principle is simply put; in an ageing (and necessarily shrinking) population, the genetic constitution of the survivors will become increasingly homogenous with advancing age as deleterious alleles contribute to disease and death and are thereby eliminated. Conversely, there is even a chance that the alleles common in the most aged may point to genes with unrecognized benefits in middle life.

In the past twenty years, it has been established that there is an association between long life and certain alleles of the major histocompatibility locus (HLA). Schächter *et al.* have looked elsewhere in the human genome, and in particular at three well characterized genes known to be involved in heart disease — those coding for the proteins apolipoprotein E, apolipoprotein B and for angiotensin-converting enzyme (ACE). What they find is mostly (but not entirely) what they expect, that the most deleterious alleles are attenuated in the centenarians.

Thus the gene for apoE is known to have three common alleles in the general population, coding for the three common variants of the protein known as apoE2, apoE3 and apoE4. The last of these has been associated with ischaemic heart disease, so that it is not surprising that its frequency among centenarians (0.052) is less than a half that in the control group. What surprises Schächter *et al.* is that the deficiency of the apoE4 allele among the centenarians is almost exactly balanced by an increase in the frequency of the allele for apoE2. Does that version of the protein confer some still obscure protec-

tive effect?

The attenuation of the frequency of the allele for apoE4 among the centenarians may not be simply the result of deaths from ischaemic heart disease. The startling question raised by Schächter *et al.* is whether the known link between this protein and late-onset Alzheimer's disease may be involved. Could the decreased frequency of the allele for apoE4 among the centenarians represent deaths among those contracting dementia in the later decades of life.

There may be a similar explanation for another puzzling finding: the frequency of the ACE allele yielding a protein with a deletion in the signal sequence is increased, despite the association of that variant of the enzyme with myocardial infarction. There are several explanations, ranging from the possibly beneficent role of the deletion-version of the enzyme in the neurons of the aged to the physical linkage of the ACE gene with that for human growth hormone, relevant to senescence in ways not yet understood.

The upshot of this intriguing study is to point to subtleties yet to be uncovered in the frequency of particular alleles at different stages of human development. It is not merely that the products of genes (ACE in this case) may have several functions, some beneficial and some deleterious, but that the balance between benefit and disadvantage may change in the course of a lifetime, which points to the complexity of the question, "How did the human genome become what it is?"

A different way of regarding that question is provided by Nicola J. Royle, Duncan M. Baird and Alec J. Jeffreys from the University of Leicester in the same issue (*Nature Genet.* 6, 52; 1994). Their purpose is to understand the differences between the human genome and those of the great apes by a search for easily recognizable nucleotide sequences near the telomeric ends of the chimpanzee chromosomes, which are known to stain differently from the corresponding regions of human chromosomes.

One outcome is a repetitive 32-bp nucleotide sequence from the region of chimpanzee chromosomes immediately inboard of the repetitive telomeric sequence TTAGGG, which is apparently common to all vertebrates. The 32-kb sequence hybridizes to DNA from chimpanzees and gorillas, but not to that from

human beings and orang-utans (not to mention both old and new world monkeys).

Moreover, the probe is found at 21 out of the 48 distinct chromosome arms of the chimpanzee genome, and is almost certainly the explanation of the distinctive G-band staining also found there. But there is no sign of hybridization with any of the human chromosomes and not, in particular, with chromosome 2, which appears to have been formed by the end-to-end fusion of two great ape chromosomes stripped of their telomeric ends.

One interpretation is that human beings are more closely related to orang-utans, both of which lack the 32-bp repetitive element, than to gorillas and chimpanzees, but that conflicts with other data. Perhaps telomeric and sub-telomeric regions of chromosomes, necessary though they are for the integrity of chromosomes, may be too unstable to be of evolutionary significance. Time, no doubt, will tell.

Meanwhile, the instability of trinucleotide repetitive DNA elements turns out to be the explanation of yet another hereditary neurological disease, dentatorubral pallidoluysian atrophy (DRPLA), so far recognized mostly in Japan. Again in the current issue, two Japanese groups from Niigata University and the National Children's Medical Research Centre in Tokyo independently report an expandable CAG repeating unit in a gene on chromosome 12 (R. Koide *et al.* & S. Nagafuchi *et al.* *Nature Genet.* 6, 7 & 14 respectively; 1994).

As with many of the other hereditary diseases with such a genetic origin identified in the past three years, the expansion of the repeating unit appears to correlate (inversely) with the age of onset of the dominant disease and (directly) with the severity of the symptoms. Again the paternal influence seems dominant, suggesting that the expansion of the triplet occurs during gametogenesis. Although DRPLA is primarily a neurodegenerative disorder marked by epilepsy and ataxia, Koide *et al.* raise the interesting question of a link with schizophrenia on the grounds that many patients have hallucinations characteristic of that disease. Interestingly, both groups have been pointed towards their identification of a gene by cDNA libraries; in neither case has the whole gene yet been defined. □

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are abbreviated; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.